A Course in Mathematical Physics 2

# Walter Thirring

# CLASSICAL FIELD THEORY

Second Edition

Springer-Verlag New York Wien

FALVEY MEMORIAL LIBRARY
VILLANOVA UNIVERSITY
VILLANOVA, PA 19085

NO LONGER PROPERTY OF
FALVEY MEMORIAL LIBRARY

Digitized by the Internet Archive
in 2022 with funding from
Kahle/Austin Foundation

https://archive.org/details/courseinmathemat0000thir

# A Course in Mathematical Physics 2

QC
20
T45 B
v.2

Walter Thirring

# A Course in Mathematical Physics

## 2

## Classical Field Theory

Second Edition

Translated by Evans M. Harrell

With 74 Illustrations

Springer-Verlag New York Wien

Dr. Walter Thirring
Institute for Theoretical Physics
University of Vienna
A-1090 Vienna
Austria

Dr. Evans M. Harrell (*Translator*)
Department of Mathematics
Georgia Institute of Technology
Atlanta, GA 30332
U.S.A.

Translation of Lehrbuch der Mathematischen Physik
Band 2: Klassische Feldtheorie
Wien-New York: Springer-Verlag 1978
© 1978 by Springer-Verlag/Wien
ISBN 3-211-81475-2 Springer-Verlag Wien New York
ISBN 0-387-81475-2 Springer-Verlag New York Wien

Library of Congress Cataloging-in-Publication Data
Thirring, Walter E.
A course in mathematical physics.
Translation of: Lehrbuch der mathematischen Physik.
Bibliography: v. 2, p.
Includes index.
Contents: — 2. Classical field theory.
1. Mathematical physics. I. Title.
QC20.T4513 vol. 2 530.1'5 86-1813

© 1979, 1986 by Springer-Verlag New York Inc.
All rights reserved. No part of this book may be translated or reproduced in any form without written permission from Springer-Verlag, 175 Fifth Avenue, New York, New York 10010, U.S.A.

Printed and bound by R. R. Donnelley & Sons, Harrisonburg, Virginia.
Printed in the United States of America.

9 8 7 6 5 4 3 2 1

ISBN 0-387-96266-2 Springer-Verlag New York Wien
ISBN 3-211-96266-2 Springer-Verlag Wien New York
ISBN 0-387-81532-5 1st ed. Springer-Verlag New York Wien
ISBN 3-211-81532-5 1. Aufl. Springer-Verlag Wien New York

# Preface

In the past decade the language and methods of modern differential geometry have been increasingly used in theoretical physics. What seemed extravagant when this book first appeared 12 years ago, as lecture notes, is now a commonplace. This fact has strengthened my belief that today students of theoretical physics have to learn that language—and the sooner the better. After all, they will be the professors of the twenty-first century and it would be absurd if they were to teach then the mathematics of the nineteenth century. Thus for this new edition I did not change the mathematical language. Apart from correcting some mistakes I have only added a section on gauge theories. In the last decade it has become evident that these theories describe fundamental interactions, and on the classical level their structure is sufficiently clear to qualify them for the minimum amount of knowledge required by a theoretician. It is with much regret that I had to refrain from incorporating the interesting developments in Kaluza–Klein theories and in cosmology, but I felt bound to my promise not to burden the students with theoretical speculations for which there is no experimental evidence.

I am indebted to many people for suggestions concerning this volume. In particular, P. Aichelburg, H. Rumpf and H. Urbantke have contributed generously to corrections and improvements. Finally, I would like to thank Dr. I. Dahl-Jensen for redoing some of the figures on the computer.

Vienna
*December, 1985*

W. Thirring

# Contents


| | | |
|---|---|---|
| | **Symbols Defined in the Text** | ix |
| **1** | **Introduction** | 1 |
| 1.1 | Physical Aspects of Field Dynamics | 1 |
| 1.2 | The Mathematical Formalism | 10 |
| 1.3 | Maxwell's and Einstein's Equations | 28 |
| **2** | **The Electromagnetic Field of a Known Charge Distribution** | 46 |
| 2.1 | The Stationary-Action Principle and Conservation Theorems | 46 |
| 2.2 | The General Solution | 57 |
| 2.3 | The Field of a Point Charge | 71 |
| 2.4 | Radiative Reaction | 89 |
| **3** | **The Field in the Presence of Conductors** | 102 |
| 3.1 | The Superconductor | 102 |
| 3.2 | The Half-Space, the Wave-Guide, and the Resonant Cavity | 113 |
| 3.3 | Diffraction at a Wedge | 125 |
| 3.4 | Diffraction at a Cylinder | 138 |
| **4** | **Gravitation** | 155 |
| 4.1 | Covariant Differentiation and the Curvature of Space | 155 |
| 4.2 | Gauge Theories and Gravitation | 174 |

4.3 Maximally Symmetric Spaces 191
4.4 Spaces with Maximally Symmetric Submanifolds 204
4.5 The Life and Death of Stars 223
4.6 The Existence of Singularities 236

**Bibliography** 253

**Index** 259

# Symbols Defined in the Text

| | | |
|---|---|---|
| $e^{i_1 i_2 \cdots i_p}$ | basis of the $p$-forms | (1.2.3) |
| $E_p(M)$ | linear space of the $p$-forms | (1.2.5; 2) |
| $d$ | exterior differential | (1.2.6) |
| $\omega_{\lvert N}$ | restriction of a form | (1.2.7; 3) |
| $E_m^0(U)$ | space of $m$-forms with compact support | (1.2.9) |
| $\langle e^i(x) \lvert e^k(x) \rangle$ | scalar product | (1.2.14) |
| $i_v$ | interior product | (1.2.16) |
| $*$ | isomorphism between $E_p$ and $E_{m-p}$ | (1.2.17) |
| $\delta$ | codifferential | (1.2.19) |
| $\Delta$ | Laplace–Beltrami operator | (1.2.20) |
| $L_v$ | Lie derivative | (1.2.23) |
| $\omega^i{}_k$ | affine connection | (1.2.25) |
| $\omega_{ik}$ | affine connection | (1.2.25) |
| $\Theta(x)$ | Heaviside step function | (1.2.31) |
| $\delta(x)$ | Dirac delta function | (1.2.31) |
| $\delta_{\hat{x}}$ | Dirac delta form | (1.2.33) |
| $G_{\hat{x}}$ | Green function | (1.2.35) |
| $\mathbf{E}$, $\mathbf{B}$, F | electric and magnetic fields | (1.3.1) |
| $A$ | vector potential | (1.3.7) |
| $\Lambda$ | gauge function | (1.3.10; 1) |
| $J$ | current | (1.3.12) |
| $Q$ | total charge | (1.3.18; 2) |
| $T^{\alpha\beta}$ | energy-momentum tensor | (1.3.20) |
| $P^\alpha$ | total energy-momentum | (1.3.21) |
| $\mathscr{T}^\alpha$ | energy-momentum form of the field | (1.3.22) |
| $z(s)$ | world-line | (1.3.25; 2) |
| $t^\alpha$ | energy-momentum form of matter | (1.3.25; 2) |
| $\mathscr{L}$ | Lagrangian | (2.1.1) |
| $W$ | action | (2.1.1) |

| | | |
|---|---|---|
| $\mathscr{S}$ | Poynting's vector | (2.1.13) |
| $D^+(N)$ | domains of influence | (2.1.15) |
| $D_{\bar{x}}$ | Green function | (2.2.5) |
| $D^{\text{ret}}(x)$ | retarded Green function | (2.2.7) |
| $G^{\text{ret}}_{\bar{x}}$ | retarded Green function (form) | (2.2.7) |
| $F^{\text{ret}}$ | retarded field strength | (2.2.9) |
| $F^{\text{in}}$ | incoming field strength | (2.2.15) |
| $F^{\text{out}}$ | outgoing field strength | (2.2.15) |
| $F^{\text{rad}}$ | radiation field | (2.2.21) |
| $D(x)$ | $D$-function | (2.2.22) |
| $\delta E$ | energy loss per period | (2.4.4; 2) |
| $j$ | specified current | (3.1.7) |
| $\varepsilon(k)$ | dielectric constant | (3.1.19; 3) |
| $S$ | superpotential | (3.1.21; 1) |
| $F(z)$ | Fresnel's integral | (3.3.10) |
| $\langle \mid \rangle$ | scalar product | (4.1.3) |
| $S_p$ | Sections | (4.1.9) |
| $D$ | exterior covariant derivative | (4.1.10) |
| $D_X$ | covariant derivative | (4.1.15) |
| $\Omega$ | curvature form | (4.1.19) |
| $R$ | curvature in space-time | (4.1.20; 1) |
| $\Gamma_{ijk}$ | Christoffel symbol | (4.1.36) |
| $R_{ijkm}$ | Riemann-Christoffel tensor | (4.1.44; 2) |
| $C_{jk}$ | Weyl forms | (4.1.44; 3) |
| $r_0$ | Schwarzschild radius | (4.4.41) |
| $K$ | curvature parameter | (4.4.42) |
| $c$ | rate of convergence | (4.6.8) |
| $J^+(x)$ | future of $x$ | (4.6.18(a)) |
| $J^-(x)$ | past of $x$ | (4.6.18(a)) |
| $C(x, S)$ | set of causal curves | (4.6.18(b)) |
| $C^1(x, S)$ | set of differentiable causal curves | (4.6.18(b)) |
| $d(\lambda)$ | length of $\lambda$ | (4.6.18(c)) |

# Introduction 1

## 1.1 Physical Aspects of Field Dynamics

*Electric and magnetic fields are dynamically interconnected in such a way that an electromagnetic disturbance propagates with a universal velocity in empty space. By studying this phenomenon we gain a qualitative understanding of field radiation and are led to expect analogous gravitational behavior.*

The unification of the theories of electric and magnetic phenomena was one of the great scientific events of the nineteenth century. Whereas stationary electric fields $\mathbf{E}$ have sources at the positions of the charges but are irrotational ($\nabla \times \mathbf{E} = \mathbf{0}$), changing magnetic fields produce circulating electromotive forces. In contrast, magnetic fields $\mathbf{B}$ are always sourceless and circulate around currents and places where there is a time-dependent electric field. The dynamical interrelation of the two fields is described by Maxwell's equations: If we consider empty space (no sources or currents), then in units where $c = 1$ they require that

$$\oint_{\partial N} d\mathbf{s} \cdot \mathbf{E} = -\int_N d\mathbf{S} \cdot \dot{\mathbf{B}}, \qquad \oint_{\partial N} d\mathbf{s} \cdot \mathbf{B} = \int_N d\mathbf{S} \cdot \dot{\mathbf{E}}, \tag{1.1.1}$$

for integrals over arbitrary surfaces $N$ with boundaries $\partial N$; and if the surface is closed, then

$$\oint_N d\mathbf{S} \cdot \mathbf{E} = \oint_N d\mathbf{S} \cdot \mathbf{B} = 0 \tag{1.1.2}$$

We shall later recognize these apparently independent relationships as different aspects of a single fact, that the field-strength form and its dual

form are closed. Before going more fully into this geometrical interpretation, let us try to come to an intuitive understanding of the physical consequences of these equations.

**Electromagnetic Waves** (1.1.3)

At a fixed time, a field

$$E_y = B_z = \cos(\omega(x - t)), \text{ all other components } 0, \tag{1.1.4}$$

looks as follows:

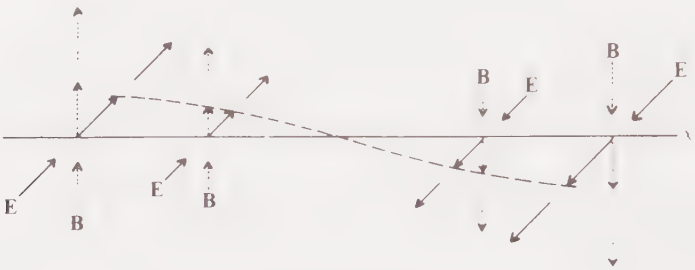



Figure 1 The fields in a plane wave

This is obviously free of sources, and it satisfies (1.1.1), as is easily seen on the surface chosen in Figure 2. Since the wave in Figure 1 moves to the right, $\dot{E}$ and $\dot{B}$ have the same sign as $E$ and $B$ in that region, and so for $t = 0$

$$\oint_{\partial N} \mathbf{E} \cdot d\mathbf{s} = -2L = -L\omega \int_0^{\pi/\omega} dx \sin \omega x = -\int_N \dot{\mathbf{B}} \cdot dS$$

We see that unlike in the stationary situation, where the electric field of a point source falls off as $1/r^2$, it is dynamically possible for it to travel through space at the speed of light, normalized here to one, without decaying, that is, without the pulse losing intensity.

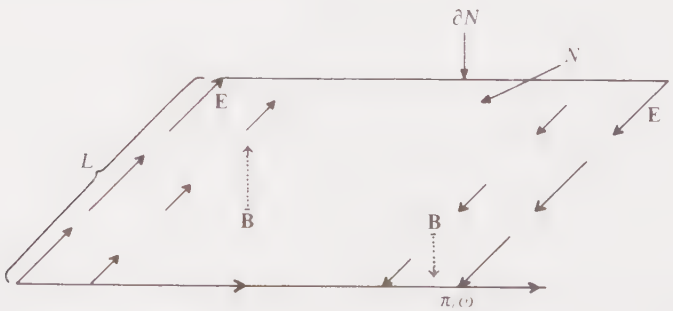



Figure 2 Illustrating the integral form of Maxwell's equations

It can also be seen from the relationships (1.1.1) and (1.1.2), though less directly, that any change in the fields propagates at the speed of light. We shall later study how this follows from the structure of the characteristics of the equivalent differential equation. For the moment let us take this property as given, and use it to investigate how an accelerated charge shakes off some of its Coulomb field and emits it as radiation.

**The Production of Electromagnetic Radiation** (1.1.5)

A charge $e$ moving with a constant velocity $\mathbf{v}$ in the $x$-direction emits no radiation (by Lorentz invariance). However, if it is brought to a stop at the origin during the time $-\tau < t < 0$, then its Coulomb field at some time $t > 0$ looks as follows: At distances $r > t + \tau$ from the origin it is equal to the field of the moving charge, since those parts of space have not yet learned of the braking of the particle. The lines of force out there point at the spot $\mathbf{x} \simeq \mathbf{v}t$ rather than at $\mathbf{x} = \mathbf{0}$, where the charge remains as $t > 0$. Thus the lines of force are displaced by $\mathbf{v}t \sim \dot{\mathbf{v}}\tau t$ compared with the field at $r < t$. At $r < t$ the field is that of a charge at rest at $\mathbf{x} = \mathbf{0}$, as the field has already forgotten that the charge ever moved. In between, in the spherical shell $t < r < t + \tau$, the lines of force progress continuously and without sources. Hence they must bend, and, as shown in Figure 3, as $r$ increases they must get folded more closely together in the parts of the spherical shell that are at most at right angles to $\mathbf{v}$. This causes the field strength to increase by a factor of $r\dot{v}$: The increase in the density of the field lines is proportional to (displacement of the field lines) (thickness of the spherical shell) $\sim \dot{v}\tau t/\tau \sim r\dot{v}$, because at $r = t \gg \tau$ one would see the field of a charge at $x = t\mathbf{v} \sim t\tau\dot{\mathbf{v}}$,

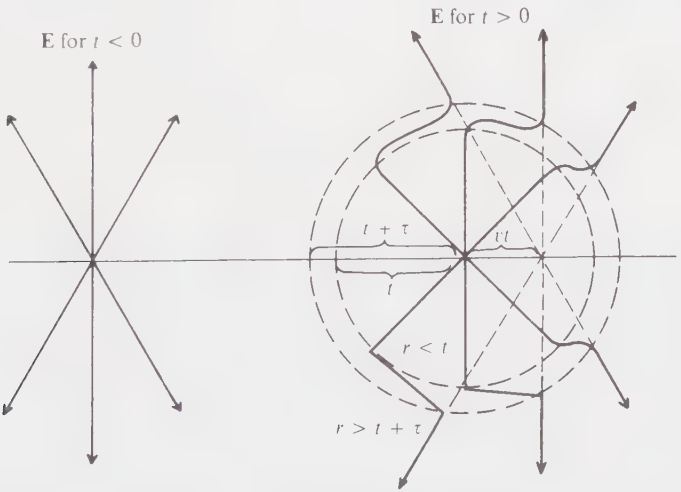



Figure 3 The field of bremsstrahlung

and thus the lines of force are displaced by this amount. The field in the spherical shell is consequently not $|\mathbf{E}| = e/r^2$, but

$$|\mathbf{E}| \sim \frac{e\dot{v}}{r}, \text{ for } t < r < t + \tau, \text{ and } \vartheta \equiv \measuredangle(x, v) \sim \frac{\pi}{2}. \tag{1.1.6}$$

**Remarks** (1.1.7)

1. The compression factor is only significant when $\vartheta$ is near $\pi$ 2. If $\vartheta = 0$ or $\pi$, then the lines of force in the spherical shell are not bent at all. A more careful calculation produces an overall factor $\sin^2 \vartheta$ in $|E|^2$.
2. The sign of the field's augmentation is clearly such that its direction is opposite to that of the acceleration.

An electric field $\sim 1/r$ rather than $\sim 1\ r^2$ as with the static Coulomb field leads immediately to the radiation of energy. The field energy contained in the spherical shell,

$$\sim \int_{t<r<t+\tau} d^3x|\mathbf{E}|^2 \sim |\mathbf{E}|^2 r^2 \tau \sim e^2|\dot{\mathbf{v}}|^2\tau$$

is conserved during the expansion, so that at large $r$ the radiated field still has just as much energy, while the Coulomb energy decreases to zero. The radiated energy is evidently imparted to the field during the braking in the time $\tau$, and then it travels off to infinity at the speed of light within the spherical shell $t < r < t + \tau$. In this way we obtain the basic formula of radiation,

**Larmor's Formula** (1.1.8)

$$\text{Radiated energy per unit time} = \frac{2}{3}\frac{e^2}{4\pi}\dot{v}^2.$$

**Remarks** (1.1.9)

1. In order to get a feel for the order of magnitude of the radiated energy, it is most convenient to use the fundamental length of electrodynamics, the classical radius of an electron, i.e., the radius at which the Coulombic potential energy equals the rest energy of an electron: $r_c = e^2\ 4\pi mc^2 \sim 10^{-13}$ cm. (As we set $c = 1$ this is equivalent to about $10^{-23}$ seconds.) If an electron at almost the speed of light is stopped in a time $\tau \sim r_c$ over a distance $b \sim \tau$, then $\dot{v} \sim v\ \tau \sim 1\ \tau \sim 1\ b$, so the braking releases a radiative energy $\sim \tau e^2\dot{v}^2 \sim e^2, b$, which is roughly the rest energy of an electron ($\sim \frac{1}{2}$ MeV) if $b \sim r_c$. Modern accelerators have made such breakneck occurrences commonplace.

2. Electrons are most easily braked with an electric field **E**, and it is natural to ask what the connection between the radiated energy and the energy of the field **E** is. Because $m\dot{\mathbf{v}} = e\mathbf{E}$, an electron radiates $E^2(e^2/m)^2\tau$, which is the fraction of the energy of the field **E** contained in a volume $\tau \times$ (the classical electron radius)$^2$. If the electron is not subjected to a single braking, but is moved periodically, as in a light wave, then we are interested in the cross-section, defined as (the radiated energy per unit time)/(the incident energy per unit time and surface area). Since the energy density of a light wave is $E^2$, the energy incident in a time $\tau$ and a unit of surface area is $E^2\tau$, and the scattering cross-section is about $r_c^2$. This means that the electron is about $10^{-13}$ cm across, in the sense that it blocks a surface area $r_c^2 \sim 10^{-26}\,\text{cm}^2$ from a beam of light. However, quantum effects often make an electron act more as if it were $10^{-11}$ cm across. An electron is best pictured as about $10^{-11}$ cm across but fairly transparent, so that it scatters light only weakly. The explanation for why matter is so often opaque, though the field is predominantly influenced by the rather transparent electrons it contains, will be discussed later.
3. The precise numerical factor given in (1.1.8) comes about because
   (a) We use units in which the Coulomb field is $e/4\pi r^2$, giving an extra $(4\pi)^{-2}$.
   (b) According to (1.1.7; 1) the energy has an angular distribution $\sim \sin^2 \vartheta$. Integrated over a spherical surface, this gives $\frac{2}{3} \cdot 4\pi$.
   (c) The energy density is actually $(|\mathbf{E}|^2 + |\mathbf{B}|^2)/2$, but the contribution from **B** equals that from **E**.

The source of light is ordinarily atoms, in which negative charges orbit positive ones, while the total charge is neutral. If a single charge oscillates, then the pattern of field lines produced is that of a repetition of one-time brakings (Figure 4, left). If two charges oscillate around each other, then this

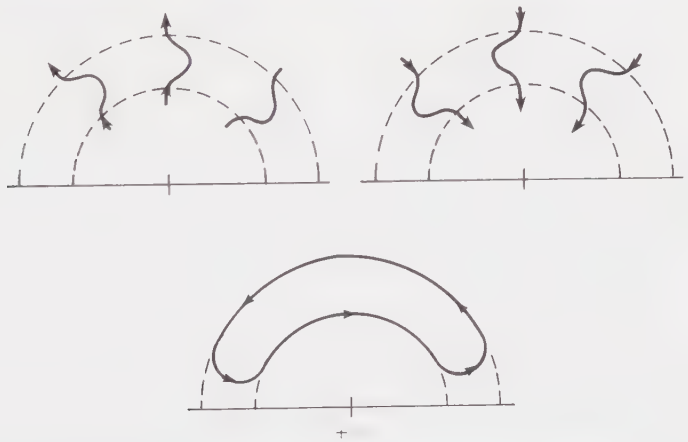


Figure 4 The electric field of an oscillating dipole

field must be superposed on one that is oppositely directed and has a phase-lag (Figure 4). As we see, oscillating dipoles emit circulating electric fields, in which $|\mathbf{E}|$ is again given by (1.1.6). If $\omega$ is the frequency and $L$ the amplitude of oscillation, then $\omega^2 L$ replaces $\dot{v}$ in (1.1.6). This produces the formula for

**Dipole Radiation** (1.1.10)

$$\text{Radiated energy per cycle} \sim e^2 L^2 \omega^3.$$

**Application to Atoms** (1.1.11)

For an atom, one would set $L$ equal to the Bohr radius $r_b \equiv \hbar^2 / me^2 = (\hbar c/e^2)^2 r_c = (137)^2 r_c \sim 10^{-8}$ cm. The electron velocity $\omega L$ is roughly $(137)^{-1}$, so the period is $\sim(137) r_b \sim (137)^3 r_c \sim 10^{-15}$ sec. Then from (1.1.10) the energy loss per cycle is $\sim(137)^{-3} e^2 / r_b$, and an electron has to orbit $(137)^3$ times to radiate an energy $e^2 / r_b$. The available energies of excitation are on the order of $e^2 / r_b \sim 10$ eV, so we expect the lifetime of an excited state to be about $(137)^3 \cdot 10^{-15}$ sec. $\sim 10^{-8}$ sec. Of course, a more exact analysis of what goes on requires a quantum-theoretical analysis of the process, but only quantum-mechanical quantity needed for a preliminary orientation is $r_b$. The creation of light may be typically outlined in this way: An atom emits $(137)^3$ waves in $10^{-8}$ seconds. Since the wavelength is about $1/\omega \sim 137 r_b$, the resultant wave-packet is $(137)^4 r_b \sim 10$ cm long (Figure 5). These figures also show up in the widths of the emitted spectral lines and in the coherence length of the radiation.

The original estimate of bremsstrahlung in (1.1.8) used only the electric field's form $\sim \mathbf{x}/r^3$. It is tempting to reason the same way with gravitation, replacing $e^2$ with the formally analogous quantity $\kappa m_p^2$. Before spending too much time on the details of this analogy, one would likely note that the numbers involved are discouragingly large. As discussed in (I: 1.1.1), the coupling

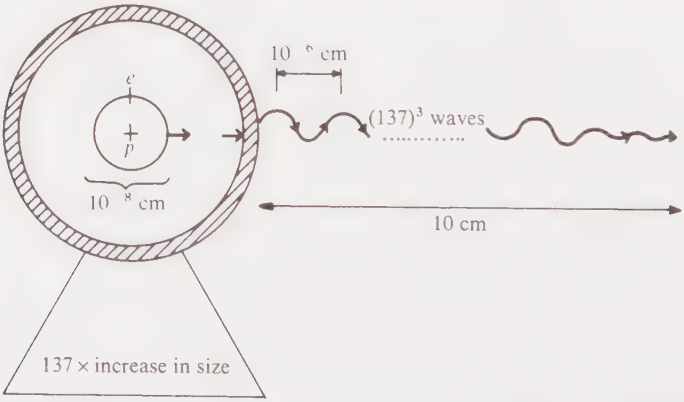



Figure 5 The typical lengths connected with the emission of light by an atom

constants differ by a factor of $10^{36}$. Whereas an atom takes $10^{-8}$ seconds to emit a photon, it would require $10^{36} \cdot 10^{-8}$ sec. $\sim 10^{28}$ sec. to bring a graviton into being. This is $10^{10}$ times the age of the universe, and it seems idle to speculate more about such questions. But masses, unlike charges, all have the same sign, and therefore large bodies benefit from tremendous coherence effects. We shall shortly see that collapsing stars ought to release gigantic energies through gravitational radiation.

**The Production of Gravitational Radiation** (1.1.12)

The points to reconsider in the derivation of (1.1.5) are:

(a) One of the facts used was that $\mathbf{E} = e\mathbf{x}\, r^{-3}$. The potential energy $e^2/r$ of two charges $e$ corresponds to the gravitational energy $\kappa M^2/r$ of two masses $M$. Hence the analogue of the Coulomb field is $\sqrt{\kappa} M\mathbf{x}\, r^{-3}$, the square of which is the energy density.

(b) The second fact used was that the electric field propagates at the speed of light. We shall later learn that this is also a property of the gravitational field, as a consequence of Einstein's equations. However, the gravitational field in turn influences the speed of light, which complicates the details of the radiation problem, though the orders of magnitude should not be affected.

(c) In the electromagnetic case the center of charge was accelerated, but the analogy fails at this point. Since the gravitational field is coupled to all masses, and the center of mass moves uniformly, this kind of radiation never occurs. This is apparent, for instance, if two masses oscillate around each other, for which two equal fields with a phase shift must be superposed, but this time with the same sign. Then in directions perpendicular to the oscillation, the large component in the direction of $\dot{\mathbf{v}}$ cancels out (see Figure 6), leaving a field $\sim 1/r^2$.

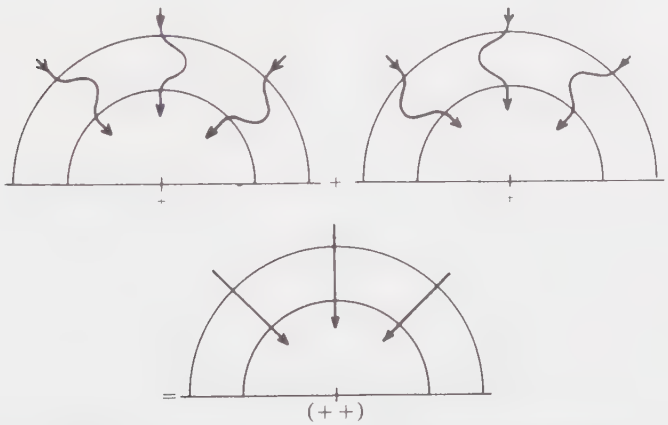



Figure 6 The gravitational field of two oscillating masses

(d) There can be electromagnetic quadrupole radiation even with a stationary center of charge; and in the situation of (c), the asymptotic fields do not quite cancel out at 45°. A more careful calculation shows that they are only reduced by a factor $v \sim \omega L$. The same thing is true for gravitational waves, leading to

**Gravitational Radiation of Rotating Masses** (1.1.13)

$$\text{Radiated energy per unit time} \sim \kappa M^2 \omega^6 L^4 \sim \frac{\kappa M^2}{L} v^5 \omega,$$

or

$$\text{Radiated energy per period} \sim v^5 \cdot \text{the gravitational energy of the system.}$$

**Remarks** (1.1.14)

1. We shall later encounter an apparently different reason for the additional factor $v^2 = \omega^2 L^2$: whereas the electric field is coupled to the current, the gravitational field interacts with the energy-momentum tensor, which is quadratic in the velocity. An additional $v$ in the coupling produces an additional $v^2$ in the radiated energy.
2. By interpreting $\kappa M^2/L$ as the gravitational energy of the system we obtain a relationship in which the miniscule coupling constant no longer appears, or, more precisely, is hidden within the gravitational energy.
3. We have not taken the trouble to puzzle out the exact numerical factor in (1.1.13). Moreover, it only applies in the context of the linear approximation to the Einsteinian theory. No exact solutions, showing how gravitational waves are created, have yet been found.

In (I: §1.1) we have discussed the qualitative features of cosmic phenomena. Let us use these rough numbers to calculate

**The Order of Magnitude of Gravitational Radiation** (1.1.15)

(a) Planets: Using Earth as an example, $v \sim 10^{-4}$, giving us a factor $10^{-20}$ in (1.1.13). That means that in Earth's $10^{10}$-year long history it has lost only one $10^{10}$-th of its potential energy through radiation, causing it to draw closer to the sun by $10^{-10}$ times the radius of its orbit, which is $\sim 10^{-10} \cdot 10^8$ km $\sim 10$ m.
(b) Double stars: For double stars $v \sim 10^{-3}$, and the orbital period is often only $10^{-3}$ years. Each year they radiate $10^{-12}$ of their gravitational energy, which was too little to be measured until quite recently.
(c) Black holes: If a star collapses under its own weight, then its contents are compressed into nuclear matter, and it turns into a neutron star,

with a radius of perhaps 10 km. (cf. §4.5). If further compressed, it produces such a strong gravitational field that the gravitational energy is comparable to the rest energy, and relativistic effects appear. The gravitational analogue of the classical radius of an electron is the Schwarzschild radius $\kappa M$, which is on the order of a kilometer for stars. Kepler orbits at this distance become relativistic, as the potential energy approaches the rest energy. The implosion of a neutron star can lead to a situation where $v \sim 1$ and $\kappa M^2/L \sim M$, and then, by (1.1.13), the radiated energy becomes equal to the rest energy of a star. Even if a more exact calculation would show that the true figure was a few percent, the energies involved are certainly huge.

Unfortunately, not much is known about gravitational radiation experimentally. It is predicted by all sensible theories, however, since our derivation has only posited a finite propagation speed for the gravitational field.

In all of the above discussion the motion of the sources of the fields has been assumed given. In fact the fields also affect the sources, and one really ought to solve the equations for the coupled system. We encountered a simple example of such a problem in the scattering of light, where an electron was accelerated by a light wave, which in turn caused the light impinging on a surface $\sim r_c^2$ to be scattered. That derivation, though, was for a free electron, and certainly does not hold for electrons in matter. Such a small cross-section would render normal matter, with interatomic distances $\sim(137)^2 r_c$, highly transparent. The reason this is not the case is that bound electrons exhibit resonance behavior, increasing $\dot{v}$ compared with (1.1.9; 2).

**The Scattering of Light by Bound Electrons** (1.1.16)

The electrons sit in the electric field $\mathbf{E}$ of the binding force as well as the light wave. Idealizing the binding as a harmonic force with strength $m\omega_0^2$, making the frequency of atomic electron oscillations $\omega_0$, would lead one to expect $m(\ddot{\mathbf{x}} + \omega_0^2 \mathbf{x}) = e\mathbf{E}$ for the equation of motion of the electrons. A periodic electric field then gives rise to an acceleration

$$\dot{\mathbf{v}} = \ddot{\mathbf{x}} \sim \frac{e\mathbf{E}}{m}\left(1 - \frac{\omega_0^2}{\omega^2}\right)^{-1},$$

which increases the scattering cross-section by $(1 - \omega_0^2/\omega^2)^{-2}$. Since the frequency $\omega$ of visible light, which is emitted by other atoms, may be near $\omega_0$, the resonance denominator can be very small, and light will not penetrate far. Of course, realistic matter has several different resonance frequencies, which is why it is colorful.

Because of the situation just sketched, it is tempting to simplify the coupled system of matter and an electromagnetic field by replacing the matter with the boundary condition that the field does not penetrate it. This is the approach usually taken in optical problems like diffraction. We shall find

that even with this simplification such problems can only be solved with difficulty, if any value is placed on precision.

If one wants to undertake a serious analysis of the coupled system of matter plus field, then it must be clearly understood that the dangers of the electric and gravitational cases lie in opposite quarters: Since like charges repel, but all masses attract, the Coulomb force tends to make a system explode, while gravity tends to make it collapse. These tendencies of the static forces are preserved in the relativistic generalizations of the equations, and can not be abolished by any mathematical sleight of hand. We shall not be able to answer the question of what holds the electronic charge together, and classical electrodynamics will remain an incomplete theory. It is replaced by quantum electrodynamics at short distances, of course, but even that theory can not explain how come the electrical energy of a point charge is finite. On the other hand, the most refined mathematics only confirms the naive expectation that, with gravity, a sufficiently massive object collapses under its own weight. It seems exceedingly difficult to get around these singularity theorems; at the very least, we can say that the prognosis of disaster is likely to be true.

## 1.2 The Mathematical Formalism

*The central concepts of classical field theory are the algebra $\{E_p\}$ of forms and their exterior derivatives. With the *-mapping $E_p \to E_{m-p}$ defined by the metric, they allow Maxwell's and Einstein's equations to be written in analogous ways.*

The calculus of *É.* Cartan is especially useful in classical field theory, because it formalizes the mathematical constructions used in the field equations. The most important rules of calculation will be briefly summarized here; the reader is referred to (I, chapter 2) or to the mathematics books cited in the bibliography for more precise definitions of the concepts.

The space of tensor fields over a manifold $M$† has a linear structure‡; every element can be written as a linear combination of certain basis elements, where the coefficients are the tensor components. In the example of the covariant vector fields, any $m$ (= the dimension of $M$ here and henceforth) linearly independent vector fields $e^i(x)$, $i = 1, \ldots, m$, $x \in M$, can be used; any vector field $V$ can be written as

$$V = \sum_{i=1}^{m} V_i(x) e^i(x).$$

† As in the first volume, a sufficiently differentiable manifold with a boundary will be referred to simply as a manifold and differentiability will likewise be implicitly assumed for tensor fields.

‡ More precisely, a module structure. The coefficients are functions of $M$ and not simply real numbers.

**Remarks** (1.2.1)

1. Independence means that the $e^i(x)$ are independent vectors for all $x$.
2. A basis is occasionally called an $m$-frame or tedrad for $m = 4$. For any $i$, $e^i$ is a vector, and not a component of a vector, or any such thing.
3. There does not generally exist a global basis, as a linearly independent set of $e^i$ can not normally be continuously extended to all of $M$. (According to (I: 2.6.17; 6), on the surface of a sphere there does not exist any continuous, nowhere vanishing vector field.) However, on the domain $U$ of a chart there is always a natural basis, as discussed below. Therefore, for local processes like differentiation there is no danger in assuming the existence of a basis.
4. A new basis can always be formed from an old one by $e^i(x) \to \bar{e}^i(x) = A^i{}_j(x)e^j(x)$, where Det $A^i{}_j(x)$ is different from zero for all $x$. No basis is distinguished from the others in the absence of more structure.

With the formation of the tensor product $\otimes$, the $e^i$ also provide a basis for the $p$-fold contravariant tensor fields. On $U$ any tensor field can be written

$$t = \sum_{j_k} t_{j_1 \cdots j_p}(x) e^{j_1} \otimes e^{j_2} \otimes \cdots \otimes e^{j_p}. \tag{1.2.2}$$

Among tensor fields, the totally antisymmetric ones ($\equiv p$-forms) are especially important, because they define $p$-dimensional volume elements. An antisymmetric tensor product $\wedge$ (the wedge, or exterior, product) can be introduced on them. Like the (finite) tensor product, it is associative and distributive, but $e^i \wedge e^j = -e^j \wedge e^i \equiv e^i \otimes e^j - e^j \otimes e^i$. Since we shall frequently use the product bases, we introduce the abbreviation

$$e^{i_1 i_2 \cdots i_p} \equiv e^{i_1} \wedge e^{i_2} \wedge \cdots \wedge e^{i_p}. \tag{1.2.3}$$

With (1.2.3) any $p$-form can be written

$$\omega = \sum_{i_k} \omega_{i_1 \cdots i_p} \frac{e^{i_1 \cdots i_p}}{p!}. \tag{1.2.4}$$

**Remarks** (1.2.5)

1. The basis $e^{i_1 \cdots i_p}$ has only $\binom{m}{p}$ independent elements, due to antisymmetry. If $p = 0$, then $p$-forms are ordinary functions, and if $p > m$ they are defined as zero, as otherwise antisymmetry would be impossible.
2. The $p$-forms on a manifold are a linear space (in fact a module), denoted by $E_p(M)$.
3. The wedge product

$$e^{i_1 \cdots i_p} \wedge e^{i_{p+1} \cdots i_{p+q}} = e^{i_1 \cdots i_{p+q}} = (-1)^{pq} e^{i_{p+1} \cdots i_{p+q}} \wedge e^{i_1 \cdots i_p}$$

creates an obviously associative mapping $E_p \times E_q \xrightarrow{\wedge} E_{p+q}$, and imparts the structure of a (graded) algebra to the space $\bigcup_p E_p$ of forms.

**The Exterior Differential** (1.2.6)

The elementary differential operations, gradient, divergence, and curl, generalize to a linear mapping $d: E_p(M) \to E_{p+1}(M)$, obeying the rules:

(a) $d(\omega_1 + \omega_2) = d\omega_1 + d\omega_2, \qquad \omega_i \in E_p$
(b) $d(\omega_1 \wedge \omega_2) = (d\omega_1) \wedge \omega_2 + (-1)^p \omega_1 \wedge d\omega_2, \qquad \omega_1 \in E_p, \omega_2 \in E_q$
(c) $d(d\omega) = 0, \qquad \omega \in E_p,$

for $p, q = 0, 1, \ldots, m$.

**Remarks** (1.2.7)

1. Rules (a) and (b) describe how $d$ acts with respect to the algebraic operations, by which Leibniz's rule gets an additional $(-1)^p$ from antisymmetry.
2. The coordinates $x^i$, $i = 1, \ldots, m$, of a point of $M$ may be thought of as a mapping $U \to \mathbb{R}$, $M \supset U =$ the domain of the chart. That is, they are an element of $E_0(U)$. The $dx^i$ are thus $m$ 1-forms, known as the **natural basis** of $E_1(U)$. If $p = 0$, then (a) and (b) are the usual rules of differentiation, so the exterior differential of a function $U \to \mathbb{R}: x \to f(x)$ becomes $df = dx^i \, \partial f / \partial x^i$. If $p = 0$, then $d$ is the gradient, and $f_{,i}$ are its components in the natural basis. Because of (c), in the natural basis

$$de^{j_1 \cdots j_p} = 0, \qquad 0 \leq p \leq m,$$

and the exterior differential of a $p$-form

$$\omega = \omega_{j_1 \cdots j_p} \frac{e^{j_1 \cdots j_p}}{p!}$$

(with the summation convention) becomes

$$d\omega = (d\omega_{j_1 \cdots j_p}) \wedge \frac{e^{j_1 \cdots j_p}}{p!} = \omega_{j_1 \cdots j_p, k} \frac{e^{k j_1 \cdots j_p}}{p!}.$$

3. Local coordinates can be introduced on any $n$-dimensional submanifold $N$ so that $x_{n+1} = x_{n+2} = \cdots = x_m = 0$. The **restriction** $\omega_{|N}$ of a form is defined by setting $dx_{n+1|N} = dx_{n+2|N} = \cdots = dx_{m|N} = 0$, and letting the restriction commute with $+$ and $\wedge$. It is then easy to see that $d\omega_{|N} = d(\omega_{|N})$.
4. Forms with vanishing exterior differentials are called **closed**, and forms that can be written as exterior differentials are called **exact**. Rule (c) means that exact $\Rightarrow$ closed. The opposite implication holds on starlike manifolds but not in general. It always holds locally.

**The Integral** (1.2.8)

Under a coordinate transformation

$$x \to \bar{x}(x), \qquad dx^i = \frac{\partial x^i}{\partial \bar{x}^j} d\bar{x}^j,$$

the natural basis of $E_m(U)$ transforms as an $m$-dimensional volume element:

$$e^{1 \quad m} = \mathrm{Det}\left(\frac{\partial x^i}{\partial \bar{x}^j}\right) \bar{e}^{1 \quad m},$$

and it is possible to define a coordinate-independent integral over $\omega$ by

$$\int_U \omega \equiv \int dx^1 \cdots dx^m \, \omega_{1 \ldots m}(x), \tag{1.2.9}$$

$$\omega \in E_m^0(U) \equiv \text{the } m\text{-forms with compact support.}$$

As $\omega \in E_m^0(M)$ can be written as a finite sum $\sum_i \omega_i$ of $m$-forms each supported in the domain of a chart, its integral is defined by (1.2.9) and $\int \sum_i \omega_i = \sum_i \int \omega_i$. More generally, the integral of $\omega \in E_n^0(M)$ over an $n$-dimensional submanifold $N$ is defined as the integral of the restriction of $\omega$ to $N$:

$$\int_N \omega \equiv \int_N \omega_{|N}. \tag{1.2.10}$$

Integration is the inverse of differentiation $d$ in the sense that Gauss's and Stokes's theorems generalize to

$$\int_N d\omega = \int_{\partial N} \omega, \qquad \omega \in E_{n-1}^0(M), \tag{1.2.11}$$

where $\partial N$ is the boundary of $N$.†

**Riemannian Structure** (1.2.12)

A symmetric, covariant tensor field of the second degree

$$g = g_{ik} e^i \otimes e^k, \text{ with } \mathrm{Det}(g_{ik}(x)) \neq 0 \quad \forall x \tag{1.2.13}$$

creates an isomorphism between the covariant and contravariant vector fields and lets the spaces of the two types of vector fields be identified. A scalar product is defined by

$$\langle e^i(x) | e^k(x) \rangle = g^{ik}(x), \qquad g^{ij} g_{jk} = \delta^i{}_k. \tag{1.2.14}$$

If the matrix $g_{ik}$ is positive, then the space is said to be **Riemannian**, and otherwise **pseudo-Riemannian**. The $e_i \equiv g_{ik} e^k$ form a dual basis, with

† We shall occasionally use this theorem when $\partial N$ is not actually a manifold, but has corners or some other harmless handicap.

$\langle e^j | e_i \rangle = \delta^j{}_i$; and $v^k \equiv g^{kl} v_l$ are called contravariant components of a vector $v = v_i e^i = v^i e_i$.

The scalar product as map $E_1 \times E_2 \to E_0$ can be generalized in two ways.

**The Scalar Product between $p$-Forms** (1.2.15)

The pseudo-Riemannian structure also induces a bilinear map $E_p \times E_p \overset{\langle | \rangle}{\to} E_0$. Expressed in components it reads

$$\langle \omega | v \rangle = \frac{1}{p!} \sum_{(i)(j)} \omega_{j_1 \dots j_p} v_{i_1 \dots i_p} g^{j_1 i_1} \cdots g^{j_p i_p} = \langle v | \omega \rangle.$$

In a Riemannian space $\langle \omega | \omega \rangle^{1/2}$ measures the $p$-dimensional volume defined by $\omega$. If it is the product of 1-forms.

**The Interior Product of a $q$-Form and a $p$-Form** (1.2.16)

A bilinear map $E_p \times E_q \to E_{p-q}$, $p \geq q$: $(\omega, v) \to i_v \omega$ can be defined by the rules:

(i) $i_v \omega = \langle \omega | v \rangle$ for $p = q$;
(ii) $i_v(\omega_1 \wedge \omega_2) = (i_v \omega_1) + (-)^{p_1} \omega_1 \wedge i_v \omega_2$ for $v \in E_1$, $\omega_i \in E_{p_i}$;
(iii) $i_{v_1 \wedge v_2} = i_{v_2} \circ i_{v_1}$.

By $\langle \, | \, \rangle$ the $E_p$ are identified with their dual spaces and the interior product is the transpose of the exterior product:

$$\langle v \wedge \omega | \mu \rangle = i_{v \wedge \omega} \mu = i_\omega (i_v \mu) = \langle \omega | i_v \mu \rangle.$$

In an $m$-dimensional manifold the space $E_m$ is one-dimensional and a pseudo-Riemannian structure distinguishes the $\varepsilon \in E_m$ which are normalized: $\langle \varepsilon | \varepsilon \rangle = (-)^s = \operatorname{Det} g \, / |\operatorname{Det} g|$. If such an $\varepsilon$ exists globally the manifold is called **orientable**. We shall assume from now on that there is globally given an orientation $\varepsilon$. In this case $E_p$ can be identified with $E_{m-p}$.

**The Hodge Duality Map** (1.2.17)

The isomorphism $E_p \overset{*}{\to} E_{m-p}$ is defined by $*\omega = i_\omega \varepsilon$. It has the properties

(i) $\varepsilon = *1$, $*\varepsilon = (-)^s$;
(ii) $* \circ * = (-)^{p(m-p)+s}$;
(iii) $i_v *\omega = *(\omega \wedge v) = (-)^{pq} i_\omega *v$ for $\omega \in E_p$, $v \in E_q$;
(iv) for $v, \omega \in E_p$ one has

$$v \wedge *\omega = \varepsilon i_v \omega = \omega \wedge *v = \varepsilon (-)^s i *v *\omega.$$

One should notice:

ad (i). $E_0$ and $E_m$ are one-dimensional and, at a point, isomorphic to $R$. $\varepsilon$ is the *-image of the number 1.

ad (ii). Except for a sign the map * is its own inverse. Unfortunately, one cannot get rid of these signs by using another $\varepsilon$.

ad (iii). Duality changes the interior product into the exterior product.

ad (iv). The word duality has its root in the fact that $E_{m-p}$ is the dual space for $E_p$ if we define a scalar product $\{\ ,\ \}$ by $v \wedge \omega = \varepsilon\{v, \omega\}$. It is related to $i$ by

$$\{v, \omega\} = (-)^{p(m-p)+s} i_v {}^*\omega.$$

So far the scalar product and other algebraic operations have been understood pointwise, that is $\langle\omega|v\rangle \in E_0 \equiv C(M)$ assigns a number to each point at a manifold. In an orientable Riemannian manifold $\int {}^*\langle\omega|v\rangle$ maps $E_p \times E_p$ into $R$ and is a scalar product in sense of a Hilbert space.

In terms of a basis the definitions imply the

**Rules** (1.2.18)

(a) $${}^*e^{i_1 \cdots i_p} = g^{i_1 j_1} \cdots g^{i_p j_p} e^{j_{p+1} \cdots j_m} \varepsilon_{j_1 \cdots j_m} \frac{\sqrt{|\mathrm{Det}\, g|}}{(m-p)!}$$

(b) $$e_j \wedge {}^*e^{j_1 \cdots j_p} = \sum_{r=1}^{p} (-1)^{r+p} \delta^{j_r}{}_j {}^*e^{j_1 \cdots j_{r-1} j_{r+1} \cdots j_p} = (-1)^{p+1} {}^*i_{e_j} e^{j_1 \cdots j_p}$$

(c) $$i_{e^j} {}^*e^{j_1 \cdots j_p} = {}^*e^{j_1 \cdots j_p j}.$$

The **codifferential**

$$\delta \equiv {}^*d^*(-1)^{m(p+1)+s}, \qquad d = {}^*\delta^*(-1)^{m(p+1)+1+s}, \tag{1.2.19}$$

is a generalization of the divergence. It is a linear mapping $E_p \to E_{p-1}$, and can be coupled with $d$ to construct the generalization $\Delta$ of the Laplace operator for $E_p \to E_p$:

$$\Delta = \delta d + d\delta \qquad \text{(the } \textbf{Laplace–Beltrami operator}\text{)}. \tag{1.2.20}$$

With a natural basis $e^{j_1 \cdots j_p}$ of $E_p(\mathbb{R}^m)$, where $g = dx^i \otimes dx^k \eta_{ik}$, $\eta_{ik} = \pm 1$ if $i = k$ and otherwise 0,

$$\begin{aligned} \delta(f e^{i_1 \cdots i_p}) &= \sum_{j-1}^{p} f^{i_j} e^{i_1 \cdots i_{j-1} i_{j+1} \cdots i_p} (-1)^{j-1}, \qquad f \in E_0(\mathbb{R}^m) \\ \Delta(f e^{i_1 \cdots i_p}) &= f_{,k}{}^k e^{i_1 \cdots i_p}, \qquad f^k_{,} \eta_{ki} \equiv f_{,i}, \end{aligned} \tag{1.2.21}$$

(Problem 6). If, in particular, $p = 1$ and $f_k$ are the components of a vector field, then $\delta$ is the ordinary divergence $f^k{}_{,k}$ and $\Delta$ is the operator $(\eta^{-1})^{ki}\, \partial^2/\partial x^i\, \partial x^k$ as applied to the individual components. (If $m = 3$ and $p = 1$, then $\Delta = -\nabla \times \nabla \times + \nabla\nabla \cdot$).

Problem 7 is to derive the following set of

**Rules** (1.2.22)

(a) $dd = \delta\delta = 0, \qquad d\Delta = \Delta d, \qquad \delta\Delta = \Delta\delta;$
(b) $\delta^* = (-1)^p * d, \qquad *\delta = (-1)^{p+1} d^*;$
(c) $d\delta^* = *\delta d, \qquad *d\delta = \delta d^*, \qquad *\Delta = \Delta^*.$

In volume I we learned moreover that if a vector field $v$ is specified, then

**The Lie Derivative** (1.2.23)

$$L_v = i_v \circ d + d \circ i_v : E_p \to E_p:$$
$$L_v(\omega_1 + \omega_2) = L_v\omega_1 + L_v\omega_2,$$
$$L_v(\omega_1 \wedge \omega_2) = (L_v\omega_1) \wedge \omega_2 + \omega_1 \wedge L_v\omega_2$$

gives the rate of change of a form under the action of the flow generated by the contravariant components of $v$. (If $v = v_i\, dx^i$ and $f \in E_0(\mathbb{R}^n)$, then $L_v f = v^i\, \partial f / \partial x^i$.)

To calculate the derivative in a different coordinate system (Problems 1 and 2) it is convenient to use a so-called orthogonal basis. Since $g_{ik}$ is a symmetric matrix, it can be transformed into a diagonal matrix $\eta_{ik}$ with only $\pm 1$ as eigenvalues, by some change of basis. This determines the $e^i$ up to a local Lorentz transformation:

$$e^i(x) \to \Lambda^i{}_k(x) e^k(x),$$
$$\eta_{kl}\Lambda^k{}_m(x)\Lambda^l{}_n(x) = \eta_{mn} \quad \forall x. \tag{1.2.24}$$

An orthogonal basis is not necessarily natural, and the exterior differentials of the orthogonal $e^i$ may not vanish. To handle the general case we define

**The Affine Connections** $\omega^i{}_k \in E_1(M)$ (1.2.25)

$$de^i = -\omega^i{}_k \wedge e^k, \qquad dg_{ik} = \omega_{ik} + \omega_{ki},$$
$$\omega_{ik} = g_{ij}\omega^j{}_k.$$

**Remarks** (1.2.26)

1. In §4.1 we shall show that these properties determine the $\omega^i{}_k$ uniquely.
2. It follows from (1.2.25) for the differentials of the bases in $E_p$ that

$$de^{j_1 \cdots j_p} = -\omega^{j_1}{}_j \wedge e^{j j_2 \cdots j_p} - \omega^{j_2}{}_j \wedge e^{j_1 j \cdots j_p} - \cdots$$
$$\cdots - \omega^{j_p}{}_j \wedge e^{j_1 \cdots j_{p-1} j},$$

and likewise

$$d{*}e^{j_1\cdots j_p} = -\omega^{j_1}{}_j \wedge {*}e^{jj_2\cdots j_p} - \omega^{j_2}{}_j \wedge {*}e^{j_1 j\cdots j_p} - \cdots$$
$$\cdots - \omega^{j_p}{}_j \wedge {*}e^{j_1\cdots j_{p-1}j}$$

(Problem 8).

3. The $\omega$'s generalize the $\Gamma$'s of (I: 1.1.7), and hence play the role of the field strength in gravitational theory. They transform inhomogeneously under a change of basis (Problem 9):

$$\omega^k{}_r \to A^k{}_s \omega^s{}_j (A^{-1})^j{}_r - (A^{-1})^j{}_r \, dA^k{}_j.$$

This means in particular that there is always a basis for which all the $\omega^k{}_r$ vanish at a point (see (I: 5.6.11)).

**Partial Differential Equations** (1.2.27)

All of the equations that will interest us are generally covariant (i.e., chart-independent), and hence the only kind of differentiation they contain is the exterior differential. Orientability and the pseudo-Riemannian structure of the world enter only through the *-mapping. The prototype of this kind of equation for a vector field is the specification of the divergence and curl of the field, or more generally of $\delta F$ and $dF$ for $F \in E_p$.

Partial differential equations determine the derivatives in certain directions, while in others they may allow the fields to vary freely, and even have discontinuities. These directions, known as the **characteristics**, are the key to an understanding of the implications of the equations. To exhibit this fact in its germinal form, we discuss the simplest nontrivial

**Example** (1.2.28)

$M = \mathbb{R}^2$ with the metrics $g_\pm \equiv dx^2 \pm dt^2$. According to (1.2.17; 2), if $p = 1$, then $* \, * = \mp \mathbf{1}$, and so * is a linear transformation of vectors, and has eigenvalues $\pm i$ for $g_+$ and $\pm 1$ for $g_-$. More specifically, if $\mathbf{F} = \mathbf{E}\, dt + \mathbf{B}\, dx$, then ${*}\mathbf{F} = -\mathbf{B}\, dt \pm \mathbf{E}\, dx$; for $g_+$, * is a rotation through 90°, and for $g_-$ it is a reflection about the axis $x = -t$ (Figure 7):

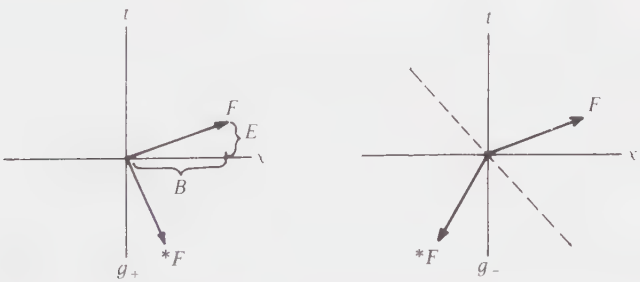



Figure 7 The action of * on $\mathbb{R}^2$

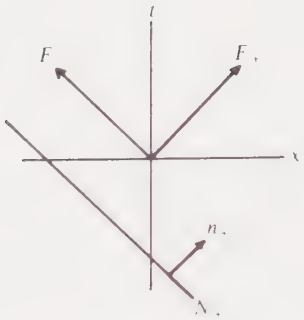



Figure 8 The hypersurface $N_+$ and its normals $n_+$

Let us therefore consider the system of equations for $\mathbf{F}$,

$$d\mathbf{F} = \mathbf{M}, \qquad \delta\mathbf{F} = \mathbf{J},$$

with specified $\mathbf{M}$ and $\mathbf{J}$. For $g$ it is possible to decouple the equations using the diagonalization of *:

$$\mathbf{F}_\pm \equiv (1 \mp *)\mathbf{F}: d\mathbf{F}_\pm = \mathbf{M} \pm *\mathbf{J}.$$

The $\mathbf{F}_\pm = (\mathbf{E} \pm \mathbf{B})(dt \pm dx)$ are independent of each other in this case, and we need only to work with a single equation. For $g_-$ the same method leads to the equations $\mathbf{F}_\pm = (1 \mp i*)\mathbf{F} = (\mathbf{E} \pm i\mathbf{B})(dt \mp i\,dx)$, in which case $\mathbf{F}_\pm$ are complex conjugates.

If $n_\pm = dx \pm dt$ and $N_\pm$ are lines parallel to $x = \mp t$, so that $n_{\pm|N_\pm} = 0$, then for $g_-$ we get $i_{n_+} n_+ = i_{n_-} n_- = 0$ and therefore

(a) $n_\pm \wedge \mathbf{F}_\pm = \mathbf{0}$;
(b) $\mathbf{F}_{\pm|N_\pm} = 0$;
(c) $i_{n_\pm} \mathbf{F}_\pm = \mathbf{0}$.

**Conclusions** (1.2.29)

(a) If $\mathbf{E} \pm \mathbf{B}$ depends only on $t \pm x$, then $d\mathbf{F}_\pm$ becomes $(\mathbf{E} \pm \mathbf{B})' n_\pm \wedge n_\pm = \mathbf{0}$. The system of equations accordingly allows $\mathbf{F}_\pm$ to vary freely in those directions; it is even possible for $\mathbf{E} + \mathbf{B}$ to be discontinuous in the direction of $n_+$. There is no other direction $n \in E_1$ such that $n \wedge \mathbf{F}_+ = \mathbf{0}$ or $n \wedge \mathbf{F}_- = \mathbf{0}$: the equations dictate how $\mathbf{F}_\pm$ varies in all the remaining directions. This state of affairs does not carry over to $g_+$, for which there are no real directions where $n \wedge \mathbf{F}_+ = \mathbf{0}$ or $n \wedge \mathbf{F}_- = \mathbf{0}$.
(b) The initial-value problem would be formulated by specifying as the initial data the restriction of $\mathbf{F}_\pm$ to some $m - 1$-dimensional submanifold $N$, in the hope that this might determine $\mathbf{F}$. If $n_{\pm|N} = 0$ at

any point of $N$, then $\mathbf{F}_{+|N}$ also vanishes at that point, making it impossible to choose the initial data at will. It is possible that they even leave $\mathbf{F}$ indeterminate. By fact (b) there exist $\mathbf{F}_+ \neq \mathbf{0}$ for which $\mathbf{F}_{+|N} = \mathbf{0}$ and $d\mathbf{F}_+ = \mathbf{0}$. If $n_{\pm|N} \neq 0$, the only condition on the initial data is $d(\mathbf{F}_{+|N}) = d\mathbf{F}_{+|N} = (\mathbf{M} - {}^*\mathbf{J})_{|N}$, which automatically vanishes (because there exists no 2-form on a one-dimensional $N$).

(c) By (1.2.23) the Lie derivative $L_{e_k}$ in the direction $x^k$ is given by $i_{e_k} \circ d + d \circ i_{e_k}$ with $e_k = g_{kj}\, dx^j$. Thus $L_{e_+}$ with $e_\pm \equiv g n_\pm = n_\mp$ is the Lie derivative in the direction $n_\pm$ (parallel to $x = \pm t$), and because $L_{e_-}\mathbf{F}_+ = i_{e_-}\, d\mathbf{F}_+ + d i_{e_-}\mathbf{F}_+ = i_{e_-}\, d\mathbf{F}_+ = i_{e_-}(\mathbf{M} - {}^*\mathbf{J})$, the rate of change of $\mathbf{F}_+$ in the direction $n_-$ is determined. As we shall see later from the explicit solution, $F$ is determined by an arbitrary set of initial data on $N$ when:

(i) $n_{\pm|N}$ is different from zero at all points of $N$; and
(ii) every line $t \pm x =$ constant intersects $N$ at exactly one point.

An $m - 1$-dimensional hypersurface $N$ (in other words, in two dimensions, a curve) that satisfies (i) and (ii) is called a **Cauchy surface**. It is nowhere tangent to the characteristics $n_\pm$, and it must be large enough to determine $\mathbf{E} \pm \mathbf{B}$ everywhere. (See Figure 9.)

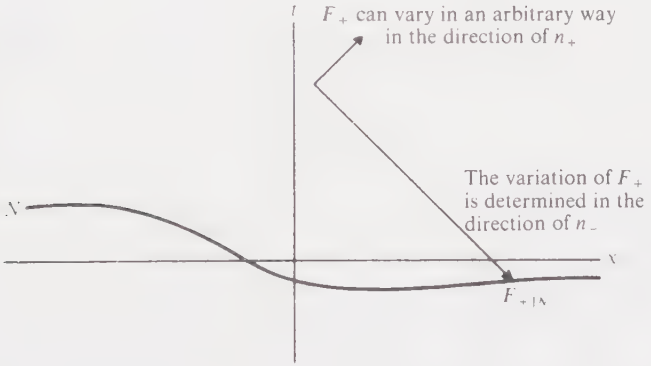



Figure 9 Determination of $F_+$ from its restrictions to the Cauchy surface $N$

**Remarks** (1.2.30)

1. Not all surfaces that are everywhere spacelike and extend to infinity are Cauchy surfaces. (See Figure 10.)
2. Observations like the foregoing can normally answer only local questions. Global questions are more difficult to treat effectively. Some seemingly innocuous manifolds, which are dear to the hearts of cosmologists, have no Cauchy surfaces at all.
3. Since $\mathbf{E} + \mathbf{B}$ can be an arbitrary function of $t + x$, one might want to risk using discontinuous functions, for which $M_\pm : n_\pm|_{M_+} = 0$ are surfaces

where $\mathbf{E}$ + $\mathbf{B}$ has jump discontinuities. If such functions are admitted, the classical notion of differentiation is fraught with unnecessary difficulties. For instance, the equation $(\partial/\partial u)(\partial/\partial v)\Phi = 0$ has the solution $\Phi = \varphi(u) + \psi(v)$, where $\varphi$ is arbitrary and $\psi$ is differentiable, but this solution does not satisfy $(\partial/\partial v)(\partial/\partial u)\Phi = 0$.

To avoid these inconveniences it is natural to try to use distributions $\varphi$, defined by their integrals against suitable test-functions. It is then always permissible to differentiate, as defined through integration by parts: $\int f\varphi^{(n)} \equiv (-1)^n \int f^{(n)}\varphi$.

A strict formulation of the problem requires the theory of locally convex topological vector spaces. As we shall not need the more profound theorems, we content ourselves with making the

**Definition** (1.2.31)

$$\Theta(x) = \begin{cases} 1, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0 \end{cases} \quad \text{(Heaviside's step function)}$$

$$\delta(x) = \frac{d\theta(x)}{dx} \qquad \text{(Dirac's delta function),}$$

and the

**Warning** (1.2.32)

Distributions form a vector space but not an algebra: they can be added but not always multiplied [22, 37].

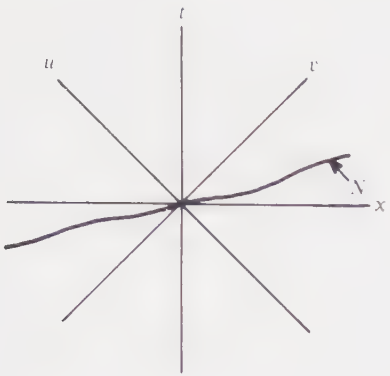
A special distribution, the Green function,† is particularly useful for solving the equations we have to deal with. This lets one exploit the linearity of the equations, the consequence of which is that the sum of the solutions for several individual sources is a solution of the equation with the sum of the sources. If the solution for a point source is known, then it can be used to construct the solution for an arbitrary inhomogeneity by superposition. Putting these ideas into practice as applied to $E_p$ requires a bit more algebra and replacing the sums with integrals.

The first thing to do is to generalize the delta function so that when it is multiplied by a $p$-form and integrated it reproduces the value of the $p$-form at a point $\bar{x} \in M$ (the set of all such values is denoted by $E_{p|\bar{x}}$). Since we are only able to integrate $m$-forms, and the result ought to be in $E_{p\,\bar{x}}$, we need a $\delta_{\bar{x}} \in E_{p|\bar{x}} \otimes E_{m-p}(M)$ such that‡

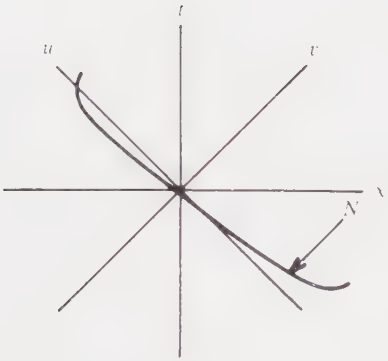
$$\int \delta_{\bar{x}} \wedge F = F(\bar{x}), \qquad \text{for all} \quad F \in E_p(M). \tag{1.2.33}$$

† We follow Jackson's [1] lead and abandon the ungrammatical phrase "the Green's function."

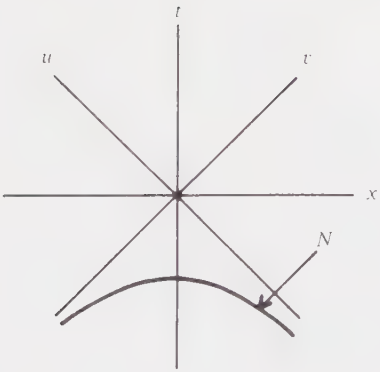
‡ In de Rham's terminology, $\delta_x$ would be called a $p$-form-valued current. In the older literature one often encounters the term Green's dyadic for the three-dimensional case. We shall write $E_p(M)$ even when the components are distributions rather than differentiable functions.




Cauchy surface, $\left.\begin{matrix} v \\ u \end{matrix}\right\} = \frac{t \pm x}{\sqrt{2}}$




not a Cauchy surface, because (i) is violated




not a Cauchy surface, because (ii) is violated

Figure 10

The distribution $\delta_{\bar{x}}$ vanishes away from $\bar{x}$ and has the form (Problem 10)

$$\delta_{\bar{x}} = \bar{e}_{i_1 \cdots i_p} \otimes {}^*e^{i_1 \cdots i_p} \frac{(-1)^{p(m-p)}}{p!} \delta(x^1 - \bar{x}^1)\delta(x^2 - \bar{x}^2) \cdots \delta(x^m - \bar{x}^m), \tag{1.2.34}$$

in a natural basis $e^{i_1 \cdots i_p}$ of a neighborhood of $\bar{x}$, and where $\bar{e}_{i_1 \cdots i_p}$ is the basis in $E_{p|x}$ and $x^i$ are the coordinates to be integrated over. The Green function $G_{\bar{x}} \in E_{p|\bar{x}} \otimes E_{m-p}(M)$ satisfies the equation

$$\Delta G_{\bar{x}} = -\delta_{\bar{x}}, \tag{1.2.35}$$

where $\Delta = d\delta + \delta d$ acts on $x^i$ with $\bar{x}$ fixed. In the cases we are interested in, we shall verify the existence of a Green function by explicit construction. Equation (1.2.35) determines $G_x$ only up to a solution of the homogeneous equation. Using $G_{\bar{x}}$, Green's formula generalizes as

$$F(\bar{x}) = (-1)^{p+m} \int_N [dG_{\bar{x}} \wedge \delta F - \delta G_{\bar{x}} \wedge dF]$$
$$- \int_{\partial N} [\delta G_{\bar{x}} \wedge F + (-1)^{m+p+s} {}^*dG_{\bar{x}} \wedge {}^*F], \tag{1.2.36}$$

which expresses $F \in E_p(M)$ at the point $\bar{x}$ in terms of $dF$ and $\delta F$ on some $m$-dimensional submanifold $N \ni \bar{x}$ and the values of the restrictions of $F$ and ${}^*F$ to $\partial N$. (For the derivation of (1.2.36), see Problem 11.)

**Example** (1.2.37)

We return to (1.2.28). $M = \mathbb{R}^2$, $p = 1$, $g = g_-$, and $(-1)^s = -1$. Using the "lightlike" coordinates that diagonalize *,

$$\begin{matrix} v \\ u \end{matrix} = \frac{t \pm x}{\sqrt{2}}, \qquad \text{for which} \quad g_- = \begin{vmatrix} 0 & -1 \\ -1 & 0 \end{vmatrix} \quad \text{and} \quad (-1)^s = -1.$$

$$*du = du, \qquad *dv = -dv, \qquad *(du \wedge dv) = -\mathbf{1},$$

and (omitting the $\otimes$ sign)

$$\delta_{\bar{x}} = (d\bar{x}\, dt - d\bar{t}\, dx)\delta(x - \bar{x})\delta(t - \bar{t})$$
$$= (d\bar{v}\, du - d\bar{u}\, dv)\delta(u - \bar{u})\delta(v - \bar{v}).$$

The Green function can be written with the aid of the step function (1.2.31) as

$$G_{\bar{x}} = (d\bar{v}\, du - d\bar{u}\, dv)\tfrac{1}{2}\Theta(\bar{u} - u)\Theta(\bar{v} - v),$$

because

$$dG_{\bar{x}} = \tfrac{1}{2}[\delta(v - \bar{v})\Theta(\bar{u} - u)d\bar{v} + \Theta(\bar{v} - v)\delta(\bar{u} - u)d\bar{u}]du \wedge dv$$
$$\delta G_{\bar{x}} = \tfrac{1}{2}[\delta(v - \bar{v})\Theta(\bar{u} - u)d\bar{v} - \Theta(\bar{v} - v)\delta(\bar{u} - u)d\bar{u}],$$

with which it is easy to verify that

$$\Delta G_{\bar{x}} = (d\bar{u}\, dv - d\bar{v}\, du)\delta(u - \bar{u})\delta(v - \bar{v}) = -\delta_{\bar{x}}.$$

If $F = \varphi\, du + \psi\, dv$, then the field equations $dF = M$, $\delta F = J$, explicitly read:

$$\begin{aligned} dF &= du \wedge dv(\psi_{,u} - \varphi_{,v}) = M \equiv du \wedge dv\, m(u, v) \\ \delta F &= \quad -\psi_{,u} - \varphi_{,v} \quad = J(u, v). \end{aligned}$$

Now the integral over $N$ of (1.2.36) contributes

$$\begin{aligned} -\int dGJ + \int \delta G M = &-\tfrac{1}{2}\int du \wedge dv(\delta(v - \bar{v})\Theta(\bar{u} - u)d\bar{v} \\ &+ \Theta(\bar{v} - v)\delta(\bar{u} - u)d\bar{u})J(u, v) \\ &+ \tfrac{1}{2}\int du \wedge dv(\delta(v - \bar{v})\Theta(\bar{u} - u)d\bar{v} \\ &- \Theta(\bar{v} - v)\delta(\bar{u} - u)d\bar{u})m(u, v) \\ = &\int_{u(\bar{v})}^{u} du\, \tfrac{1}{2}(m(u, \bar{v}) - J(u, \bar{v}))d\bar{v} \\ &- \int_{v(\bar{u})}^{v} dv\, \tfrac{1}{2}(m(\bar{u}, v) + J(\bar{u}, v))d\bar{u}, \end{aligned}$$

if the submanifold $\partial N$ is defined as $\{u = u(v)\} = \{v = v(u)\}$:

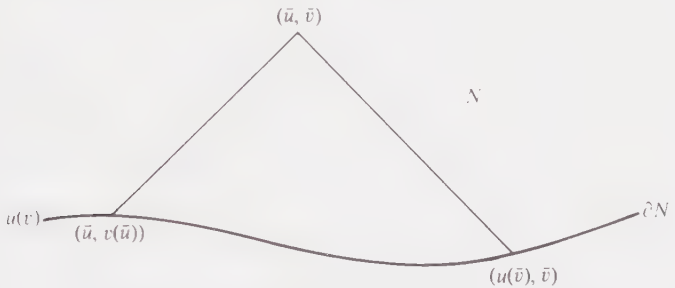



Figure 11 Variables for the fields at a boundary

This contribution clearly satisfies $2\psi_{,u} = m - J$ and $2\varphi_{,v} = -m - J$, and approaches zero as $(\bar{u}, \bar{v})$ approaches $\partial N$. To calculate the boundary term $F_{\partial N} \equiv \int_{\partial N} \cdots$ note that $\partial N$ has the orientation of $-dx$ for this boundary integral (see (I: 2.6.8; 1)):

$$\begin{aligned} \int_{\partial N} \delta G \wedge F = &\frac{d\bar{v}}{2}\int_{\partial N} (\varphi\, du + \psi\, dv)\delta(v - \bar{v})\Theta(\bar{u} - u) \\ &- \frac{d\bar{u}}{2}\int_{\partial N} (\varphi\, du + \psi\, dv)\delta(u - \bar{u})\Theta(\bar{v} - v) \\ \int_{\partial N} {*dG} \wedge {*F} = &\frac{d\bar{v}}{2}\int_{\partial N} (-\varphi\, du + \psi\, dv)\delta(v - \bar{v})\Theta(\bar{u} - u) \\ &+ \frac{d\bar{u}}{2}\int_{\partial N} (-\varphi\, du + \psi\, dv)\delta(u - \bar{u})\Theta(\bar{v} - v); \end{aligned}$$

since $dv$ contributes positively and $du$ negatively, the net result is that

$$F_{\partial N} = d\bar{v}\, \psi(u(\bar{v}), \bar{v})\Theta(\bar{u} - u(\bar{v})) + d\bar{u}\, \varphi(\bar{u}, v(\bar{u}))\Theta(\bar{v} - v(\bar{u})).$$

As long as $\bar{x} \in N$, so that $\Theta(\bar{u} - u(\bar{v}))\Theta(\bar{v} - v(\bar{u})) = \mathbf{1}$, this solves the homogeneous equation, because $\psi$ is independent of $\bar{u}$ and $\varphi$ is independent of $\bar{v}$. It also clearly satisfies the boundary conditions.

**Remarks** (1.2.38)

1. It was not apparent in (1.2.36) that $\partial N$ was required to be a Cauchy surface. However, the example shows that $dG_{\bar{x}}$ can not be used as a distribution on $\partial N$ if $\partial N$ contains parts of the light-cone of $\bar{x}$, which would contribute $\delta(0)$.
2. The derivation relies on Stokes's theorem, which we only know for forms of compact support. Since the support of $G_{\bar{x}}$ remains on a light-cone, Stokes's theorem may also be used here for non-compact $N$.
3. At this point (1.2.36) appears as an identity for a solution of the field equations. It may be used to construct a solution from its initial data, but gives no information about how arbitrarily the initial data may be chosen. In the example, if $\partial N$ is a Cauchy surface, then the solution meets all requirements with arbitrary $\psi(u(v), v)$ and $\varphi(u, v(u))$ on $\partial N$. The initial values can be specified independently, since $\partial N$ never contains two points of the same light-cone, while information about an initial condition only propagates along its characteristics. If the metric were $g_-$, then the situation would be quite different. The solutions would be analytic functions of $x + it$, and would be determined everywhere by their values on an arbitrarily small curve segment.
4. The equations of constraint mentioned in (1.2.29(b)) arise in 4 dimensions.

In particular, Formula (1.2.36) solves Maxwell's equations, and electrodynamics consists largely in working out special cases of it. The normal detour via the introduction of potentials is inconvenient as well as unnecessary, as one really wants to express $F$ in terms of the boundary values of $F$ and $^*F$, and not of the potentials. Unfortunately, the canonical formulation of the equations of motion, and hence also quantum mechanics, use potentials. Yet even the potentials can be obtained from (1.2.36). On the other hand, the currents analogous to $J$ and $M$ appearing in Einstein's equations depend nonlinearly on the fields, so in that case (1.2.36) is useful only in the linear approximation.

**Problems** (1.2.39)

1. The metric $g = d\rho^2 + \rho^2\, d\varphi^2 + dz^2$ in cylindrical coordinates for $\mathbb{R}^3 \backslash \{\mathbf{0} \times \mathbb{R}\}$. Calculate $\omega_{ik}$ for the orthogonal basis $e^1 = dz, e^2 = d\rho, e^3 = \rho\, d\varphi$. Write $\nabla \times A$ and

$\nabla \cdot A$ in the components of this basis and of the natural basis. What is the connection between them?

2. The same question for spherical coordinates and $\mathbb{R}^3\backslash\{\mathbf{0} \times \mathbb{R}\}$, for which $g = dr^2 + r^2\, d\vartheta^2 + r^2 \sin^2 \vartheta\, d\varphi^2$.

3. Calculate $\Delta$ from (1.2.20) for $p = 0$ in the natural basis, and specialize to the cases of cylindrical and spherical coordinates on $\mathbb{R}^3\backslash\{\mathbf{0} \times \mathbb{R}\}$.

4. Prove the normalization $* \circ * = (-1)^{p(m-p)+s}1$. Using (1.2.16).

5. Derive the rules (1.2.18):

$$*e^{i_1 \cdots i_p} = g^{i_1 j_1} \cdots g^{i_p j_p} e^{j_{p+1} \cdots j_m} \varepsilon_{j_1 \cdots j_m} \frac{\sqrt{|g|}}{(m-p)!},$$

$$e_j \wedge *e^{j_1 \cdots j_p} = \sum_r (-1)^{r+p} \delta^{j_r}{}_j *e^{j_1 \cdots j_{r-1} j_{r+1} \cdots j_p}$$

$$i_{e_j} *e^{j_1 \cdots j_p} = *e^{j_1 \cdots j_p j}.$$

6. Verify (1.2.21).

7. Verify (1.2.22).

8. Check (1.2.26; 2).

9. Derive the transformation law for the $\omega$'s (1.2.26; 3).

10. Show that $\delta_{\bar{x}}$ from (1.2.34) has the property (1.2.33).

11. Prove (1.2.36).

12. Find a manifold $M$ and a $J \in E_p(M)$ for which the equation $dF = J$ has no (global) solution, although $dJ = 0$.

**Solutions** (1.2.40)

1. Orthogonal basis: $\eta_{ik} = \delta_{ik}$, $0 = de^1 = de^2$, $de^3 = e^{23}/\rho$. Hence only $\omega_{32} = -\omega_{23} = e^3/\rho \neq 0$, and $de^{12} = de^{32} = 0$, $de^{31} = -\omega^{32} \wedge e^{21}$. Letting $A = a_i e^i$,

$$dA = (a_{1,2} - a_{2,1})e^{21} + (a_{2,3}/\rho - a_{3,2} - a_3/\rho)e^{32} + (a_{3,1} - a_{1,3}/\rho)e^{13},$$

$$*dA = (a_{3,2} + a_3/\rho - a_{2,3}/\rho)e^1 + (a_{1,3}/\rho - a_{3,1})e^2 + (a_{2,1} - a_{1,2})e^3,$$

$$*A = a_1 e^{23} + a_2 e^{31} + a_3 e^{12},\ *d*A = a_{1,1} + a_{2,2} + a_{3,3}/\rho + a_2/\rho.$$

Natural basis: $\sqrt{|g|} = \rho$. Let $A = A_z\, dz + A_\rho\, d\rho + A_\varphi\, d\varphi$. Then

$$*A = \rho[A_z\, d\rho \wedge d\varphi + A_\rho\, d\varphi \wedge dz + \rho^{-2} A_\varphi\, dz \wedge d\rho],$$

$$*d*A = \rho^{-1}[(\rho A_z)_{,z} + (\rho A_\rho)_{,\rho} + (\rho^{-1} A_\varphi)_{,\varphi}],$$

$$*dA = \rho^{-1}[(A_{\varphi,\rho} - A_{\rho,\varphi})dz + (A_{z,\varphi} - A_{\varphi,z})d\rho + \rho^2(A_{\rho,z} - A_{z,\rho})d\varphi].$$

The connection is that $(A_z, A_\rho, A_\varphi) = (a_1, a_2, \rho a_3)$.

2. Orthogonal basis: $e^1 = dr$, $e^2 = r\,d\vartheta$, $e^3 = r\sin\vartheta\,d\varphi$, $de^1 = 0$, $de^2 = dr \wedge d\vartheta$, $de^3 = r\cos\vartheta\,d\vartheta \wedge d\varphi + \sin\vartheta\,dr \wedge d\varphi \Rightarrow \omega_{21} = e^2/r$, $\omega_{31} = \sin\vartheta\,d\varphi$, $\omega_{32} = \cos\vartheta\,d\varphi \Rightarrow de^{12} = 0$, $de^{13} = -e^{23}/r\tan\vartheta$, $de^{23} = 2e^{23}/r$. Let $A = a_i e^i$. Then

$$*d*A = a_{1,1} + \frac{2}{r}a_1 + \frac{1}{r}a_{2,2} + \frac{a_2}{r\tan\vartheta} + \frac{a_{3,3}}{r\sin\vartheta},$$

$$*dA = \left[\frac{a_{3,2}}{r} + \frac{a_3}{r\tan\vartheta} - \frac{a_{2,3}}{r\sin\vartheta}\right]e^1 + \left[\frac{a_{1,3}}{r\sin\vartheta} - \frac{a_3}{r} - a_{3,1}\right]e^2 + \left(a_{2,1} + \frac{a_2}{r} - \frac{a_{1,2}}{r}\right)e^3.$$

Natural basis: $\sqrt{|g|} = r^2\sin\vartheta$. Let $A = A_r\,dr + A_\vartheta\,d\vartheta + A_\varphi\,d\varphi$. Then

$$*A = r^2\sin\vartheta[A_r\,d\vartheta \wedge d\varphi + r^{-2}A_\vartheta\,d\varphi \wedge dr + r^{-2}\sin^{-2}\vartheta A_\varphi\,dr \wedge d\vartheta]$$

$$*dA = \frac{dr}{r^2\sin\vartheta}(A_{\varphi,\vartheta} - A_{\vartheta,\varphi}) + \frac{d\vartheta}{\sin\vartheta}(A_{r,\varphi} - A_{\varphi,r}) + d\varphi\sin\vartheta(A_{\vartheta,r} - A_{r,\vartheta}),$$

$$*d*A = \frac{1}{r^2\sin\vartheta}\left[(r^2\sin\vartheta A_r)_{,r} + (\sin\vartheta A_\vartheta)_{,\vartheta} + \left(\frac{A_\varphi}{\sin\vartheta}\right)_{,\varphi}\right].$$

The connection is that $(A_r, A_\vartheta, A_\varphi) = (a_1, ra_2, r\sin\vartheta a_3)$.

3. $$\Delta f = \frac{1}{\sqrt{g}}\frac{\partial}{\partial x^\alpha}(f_{,\beta}g^{\alpha\beta}\sqrt{g})$$

Cylindrical coordinates: $\Delta = \frac{1}{\rho}\left[\rho\frac{\partial^2}{\partial z^2} + \frac{\partial}{\partial\rho}\rho\frac{\partial}{\partial\rho} + \frac{1}{\rho}\frac{\partial^2}{\partial\varphi^2}\right]$

Spherical coordinates: $\Delta = \frac{1}{r^2\sin\vartheta}\left[\frac{\partial}{\partial r}r^2\sin\vartheta\frac{\partial}{\partial r} + \frac{\partial}{\partial\vartheta}\sin\vartheta\frac{\partial}{\partial\vartheta} + \frac{\partial}{\partial\varphi}\frac{1}{\sin\vartheta}\frac{\partial}{\partial\varphi}\right]$.

4. From (1.2.16) we infer

$$\langle *\omega|*\omega\rangle = \langle *\omega|i_\omega\varepsilon\rangle = \langle\omega \wedge *\omega|\varepsilon\rangle$$
$$= (-)^{p(m-p)}\langle *\omega \wedge \omega|\varepsilon\rangle = (-)^{p(m-p)}\langle\omega|i_{*\omega}\varepsilon\rangle = (-)^{p(m-p)}\langle\omega|**\omega\rangle.$$

The fact that $**\omega$ can differ from $\omega$ only by a sign is a general property of duality. Therefore $**\omega = (-)^{p(m-p)+s}\omega$ is equivalent to $\langle\omega|\omega\rangle = (-)^s\langle *\omega|*\omega\rangle$. It is obviously enough to show this for a basis

$$e^{i_1\ldots i_p} \equiv \varepsilon_p \in E_p$$

such that $\varepsilon = \varepsilon_p \wedge \varepsilon_{m-p}$ with $i_{\varepsilon_p}\varepsilon_{m-p} = 0$. Then

$$\langle *\varepsilon_p|*\varepsilon_p\rangle = \langle\varepsilon_{m-p}|\varepsilon_{m-p}\rangle\langle\varepsilon_p|\varepsilon_p\rangle\langle\varepsilon_p|\varepsilon_p\rangle$$
$$= \langle\varepsilon|\varepsilon\rangle\langle\varepsilon_p|\varepsilon_p\rangle = (-)^s\langle\varepsilon_p|\varepsilon_p\rangle.$$

5. In a basis we have

$$\varepsilon = \frac{e^{i_1\ldots i_m}}{m!}\varepsilon_{i_1\ldots i_m}\sqrt{|g|}.$$

where

$$g = \mathrm{Det}\, g_{ik} = (\mathrm{Det}\langle e^i | e^k \rangle)^{-1}.$$

Then (a) follows from the definition $*\omega = i_\omega \varepsilon$ and the rules (1.2.16) for $i$. (b) and (c) are a consequence of (1.2.17(iii)).

6.

$$\begin{aligned}
\delta(f e^{i_1 \cdots i_p}) &= *[(-1)^{m(p+1)+s} f_{,k} e^k \wedge *e^{i_1 \cdots i_p}] \\
&= *\left[(-1)^{(m+1)(p+1)+s} \sum_{j=1}^{p} f^{,i_j} * e^{i_1 \cdots i_{j-1} i_{j+1} \cdots i_p} (-1)^{j+1}\right] \\
&= \sum_{j=1}^{p} f^{,i_j} e^{i_1 \cdots i_{j-1} i_{j+1} \cdots i_p} (-1)^{j+1}.
\end{aligned}$$

$$d\delta(f e^{i_1 \cdots i_p}) = \sum_{j=1}^{p} f^{,i_j}_{,i_0} e^{i_0 i_1 \cdots i_{j-1} i_{j+1} \cdots i_p} (-1)^{j-1}$$

$$\delta d(f e^{i_1 \cdots i_p}) = \sum_{j=0}^{p} f^{,i_j}_{,i_0} e^{i_0 i_1 \cdots i_{j-1} i_{j+1} \cdots i_p} (-1)^{j}$$

$$\Rightarrow \Delta(f e^{i_1 \cdots i_p}) = \sum_{k=1}^{m} f^{,k}_{,k} e^{i_1 \cdots i_p}.$$

7. (a) $d\Delta = d\delta\, d = \Delta d$, $\delta\Delta = \delta\, d\delta = \Delta\delta$;
(b) $\delta^* = *d**(-1)^{m(m-p)+1+s} = (-1)^p * d$, $d^* = *\delta**(-1)^{m(m-p)+1+s} = (-1)^{p+1} * \delta$;
(c) $d\delta^* = (-1)^p d^* d = *\delta d$, $\delta d^* = (-1)^{p+1} \delta^* \delta = *d\delta$.

8. Letting $\omega_k{}^l = \eta^{lj}\omega_{kj}$, we start with the identity

$$\omega_{k_1}{}^i \varepsilon_{i k_2 \cdots k_m} + \omega_{k_2}{}^i \varepsilon_{k_1 i \cdots k_m} + \cdots + \omega_{k_m}{}^i \varepsilon_{k_1 \cdots k_{m-1} i} = 0$$

$\forall k_1 = 1, \ldots, m; \ldots; k_m = 1, \ldots, m$. To verify this, consider the three cases:

(i) All $k_i$ are different. Then $i$ must equal the $k$ that is missing in $\varepsilon$, and $\omega_{kj} = -\omega_{jk}$; since $\eta$ is diagonal, $\omega_k{}^i = 0$ if $i = k$.
(ii) Two of the $k_i$ are equal, say $k_1 = k_2$. There remains

$$\omega_{k_1}{}^i \varepsilon_{i k_1 \ldots} + \omega_{k_1}{}^i \varepsilon_{k_1 i \ldots} = 0.$$

(iii) Three $k_i$ are equal. Then all the $\varepsilon$'s vanish.

If the identity is multiplied by $e^{k_{p+1} \cdots k_m}$ and the indices are relabeled, then

$$\omega_{k_{p+1}}{}^i \varepsilon_{k_1 \cdots k_p i k_{p+2} \cdots m} e^{k_{p+1} \cdots k_m} = -\omega^{k_{p+1}}{}_i \varepsilon_{k_1 \cdots k_m} e^{i k_{p+2} \cdots k_m},$$

in the orthogonal basis. In general, because of (1.2.18(a)),

$$\begin{aligned}
(m-p)! d^* e^{j_1 \cdots j_p} &= \eta^{j_1 k_1} \cdots \eta^{j_p k_p} \varepsilon_{k_1 \cdots k_m} d e^{k_{p+1} \cdots k_m} \\
&= -\eta^{j_1 k_1} \cdots \eta^{j_p k_p} \varepsilon_{k_1 \cdots k_m} \{\omega^{k_{p+1}}{}_i e^{i k_{p+2} \cdots k_m} + \cdots + \omega^{k_m}{}_i e^{k_{p+1} \cdots k_{m-1} i}\} \\
&= \eta^{j_1 k_1} \cdots \eta^{j_p k_p} \{\omega_{k_{p+1}}{}^i \varepsilon_{k_1 \cdots k_p i k_{p+1} \cdots k_m} + \omega_{k_{p+2}}{}^i \varepsilon_{k_1 \cdots k_{p+1} i k_{p+3} \cdots k_m} + \cdots \\
&\quad + \omega_{k_m}{}^i \varepsilon_{k_1 \cdots k_{m-1} i}\} e^{k_{p+1} k_{p+2} \cdots k_m} \\
&= -\eta^{j_1 k_1} \cdots \eta^{j_p k_p} \{\omega_{k_1}{}^i \varepsilon_{i k_2 \cdots k_m} + \cdots + \omega_{k_p}{}^i \varepsilon_{k_1 \cdots k_{p-1} i k_{p+1} \cdots k_m}\} e^{k_{p+1} \cdots k_m} \\
&= -\omega^{j_1}{}_i * e^{i j_2 \cdots j_p} - \cdots - \omega^{j_p}{}_i * e^{j_1 j_2 \cdots j_{p-1} i} (m-p)!.
\end{aligned}$$

9. Let

$$\bar{e}^j = A^j{}_k e^k : d\bar{e}^j = dA^j{}_r (A^{-1})^r{}_k \bar{e}^k - A^j{}_k \omega^k{}_r (A^{-1})^r{}_s \bar{e}^s = -\bar{\omega}^j{}_k \bar{e}^k$$

$$\Rightarrow \bar{\omega}^j{}_k = A^j{}_s \omega^s{}_r (A^{-1})^r{}_k - (A^{-1})^s{}_k \, dA^j{}_s,$$

because this also satisfies the second defining equation:

$$\bar{g} = A^{-1\prime} g A^{-1} \Rightarrow d\bar{g} = A^{-1\prime}(dg)A^{-1} - A^{-1\prime} g A^{-1}(dA)A^{-1}$$
$$- A^{-1\prime}(dA')A^{-1\prime} g A^{-1} = \bar{g}\bar{\omega} + (\bar{g}\bar{\omega})'$$
$$= A^{-1\prime} g \omega A^{-1} - A^{-1\prime} g A^{-1}(dA)A^{-1} + A^{-1\prime}(g\omega)' A^{-1} - A^{-1\prime}(g A^{-1} dA)' A^{-1}.$$

10. According to (1.2.18),

$$\frac{1}{p!} e_{j_1 \cdots j_p} \wedge {}^*e^{i_1 \cdots i_p} \omega^{j_1 \cdots j_p} = \omega^{i_1 \cdots i_p} {}^*\mathbf{1}.$$

Thus

$$\bar{e}_{i_1 \cdots i_p} \otimes {}^*e^{i_1 \cdots i_p} \wedge e_{j_1 \cdots j_p} \omega^{j_1 \cdots j_p} \frac{(-1)^{p(m-p)}}{p!} = \bar{e}^{i_1 \cdots i_p} \omega_{i_1 \cdots i_p} {}^*\mathbf{1}.$$

Integrating this $\int dx_1 \cdots dx_m$ yields

$$\int \delta_{\bar{x}} \wedge e^{i_1 \cdots i_p} \omega_{i_1 \cdots i_p} = \bar{e}^{i_1 \cdots i_p} \omega_{i_1 \cdots i_p}(\bar{x}).$$

11.

$$\delta G_{\bar{x}} \wedge dF = (-1)^{p+m+1}[d(\delta G_{\bar{x}} \wedge F) - (d\delta G_{\bar{x}}) \wedge F],$$
$$dG_{\bar{x}} \wedge \delta F = (-1)^{mp+m+s} d{}^*F \wedge {}^*dG_{\bar{x}}$$
$$= (-1)^{mp+m+s}[d(*F \wedge {}^*dG_{\bar{x}} - (-1)^{m-p} {}^*F \wedge d{}^*dG_{\bar{x}}]$$
$$= (-1)^{mp+m+s} d(*F \wedge {}^*dG_{\bar{x}}) + (-1)^{mp+p+s+1} {}^*d{}^*dG_{\bar{x}} \wedge F$$
$$= (-1)^{mp+m+s} d(*F \wedge {}^*dG_{\bar{x}}) + (-1)^{m+1+p} \delta dG_{\bar{x}} \wedge F$$
$$\Rightarrow (-1)^{p+m}[dG_{\bar{x}} \wedge \delta F - \delta G_{\bar{x}} \wedge dF]$$
$$= \underbrace{-(\delta d + d\delta)G_{\bar{x}}}_{=\delta_{\bar{x}}} \wedge F + d(\delta G_{\bar{x}} \wedge F + (-1)^{mp+p+s} {}^*F \wedge {}^*dG_{\bar{x}}),$$

and when this is integrated over $N$, we get (1.2.36).

12. $M = T^1$ and $J = d\varphi \in E_1(T^1)$, where $\varphi$ is the angle on the torus. If $F \in E_0(T^1)$, then it would be $\varphi +$ constant, which can not be defined continuously over all of $T^1$, even though $d\varphi$ exists on all of $T^1$.

## 1.3 Maxwell's and Einstein's Equations

*Field strengths are described by 2-forms, the exterior differentials of which in the electric (respectively gravitational) case are the 3-forms of charge (resp. energy and momentum). These 3-forms, which function as the sources of the fields, are exact, a fact that implies differential and integral conservation theorems.*

Maxwell wrote Faraday's discoveries in the form of the equations that now bear his name (Table 1), and which describe all electromagnetic phenomena. The way they fuse space and time, and the electric and magnetic fields, is so awe-inspiring that Boltzmann, quoting Goethe, exclaimed, "Was it a god who traced these signs?"† It is regrettable that Cartan's calculus did not exist at that time, as it has perfected the notation in which all the facts of electromagnetism are so concisely formulated.

As discussed in detail in (I: §5.1), the electric and magnetic fields **E** and **B** are combined in a single 2-form $F$, which is written in the natural basis of $\mathbb{R}^4$, with coordinates $x^0 = t$, $\mathbf{x} = (x^1, x^2, x^3)$ and in vector notation, as

$$F = \tfrac{1}{2}F_{\alpha\beta}\, dx^\alpha \wedge dx^\beta = dt \wedge (d\mathbf{x} \cdot \mathbf{E}) - dx^1 \wedge dx^2 B_3 - dx^3 \wedge dx^1 B_2 - dx^2 \wedge dx^3 B_1. \quad (1.3.1)$$

**Table 1**

**Maxwell's Equations in the Course of History**

The constants $c$, $\mu_0$, and $\varepsilon_0$ are set to 1, and modern notation is used for the components.

| The Homogeneous Equation | The Inhomogeneous Equation |
|---|---|
| Earliest Form | |
| $\frac{\partial B_x}{\partial x} + \frac{\partial B_y}{\partial y} + \frac{\partial B_z}{\partial z} = 0$ | $\frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} = \rho$ |
| $\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} = -\dot{B}_x$ | $\frac{\partial B_z}{\partial y} - \frac{\partial B_y}{\partial z} = j_x + \dot{E}_x$ |
| $\frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} = -\dot{B}_y$ | $\frac{\partial B_x}{\partial z} - \frac{\partial B_z}{\partial x} = j_y + \dot{E}_y$ |
| $\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} = -\dot{B}_z$ | $\frac{\partial B_y}{\partial x} - \frac{\partial B_x}{\partial y} = j_z + \dot{E}_z$ |
| At the End of the Last Century | |
| $\mathbf{V} \cdot \mathbf{B} = 0$ | $\mathbf{V} \cdot \mathbf{E} = \rho$ |
| $\mathbf{V} \times \mathbf{E} = -\dot{\mathbf{B}}$ | $\mathbf{V} \times \mathbf{B} = \mathbf{j} + \dot{\mathbf{E}}$ |
| At the Beginning of This Century | |
| $*F^{\beta\alpha}{}_{,\alpha} = 0$ | $F^{\beta\alpha}{}_{,\alpha} = j^\beta$ |
| Mid-Twentieth Century | |
| $dF = 0$ | $\delta F = J$ |

† Faust, part I, Dr. Faust's first soliloquy.

**The Homogeneous Maxwell Equations** (1.3.2)

These are the basis-free statement that $F$ is closed:

$$dF = 0. \tag{1.3.3}$$

By (1.2.11) this is equivalent to

$$0 = \int_{\partial N_3} F, \tag{1.3.4}$$

if $F_{|N_3} \in E_2^0(N_3)$, and $N_3$ is three-dimensional. To illustrate this concentrated notation, let us examine some

**Special Cases** (1.3.5)

1. $N_3 = \{(t, \mathbf{x}) \in \mathbb{R}^4 : t = t_0 \in \mathbb{R}, |\mathbf{x}| \leq R\}$ (a spherical ball)
   $\partial N = \{(t, \mathbf{x}) \in \mathbb{R}^4 : t = t_0, |\mathbf{x}| = R\}$ (a spherical surface). (See Figure 12.)

   Then (1.3.4) says that there are no magnetic charges:

$$\int_{|\mathbf{x}| = R} \mathbf{B} \cdot d\mathbf{S} = 0.$$

   The differential version of this statement, $\nabla \cdot \mathbf{B} = 0$, amounts to $d(F_{|N_3}) = 0$, which follows from (1.3.3) because $d$ commutes with restrictions.

2. $N_3 = \{(t, \mathbf{x}) \in \mathbb{R}^4 : t_0 \leq t \leq t_1, x^3 = 0, |\mathbf{x}| \leq R\}$ (a cylinder).
   $\partial N = \{(t, \mathbf{x}) \in \mathbb{R}^4 : t_0 \leq t \leq t_1, x^3 = 0, |\mathbf{x}| = R\} \cup \{(t, \mathbf{x}) \in \mathbb{R}^4 :$
   $t = t_0, x^3 = 0, |\mathbf{x}| \leq R\} \cup \{(t, \mathbf{x}) : t = t_1, x^3 = 0, |\mathbf{x}| \leq R\}$

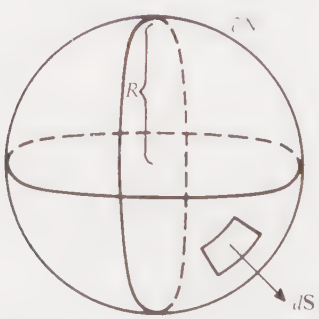



Figure 12 $N_3$ of (1.3.5; 1)

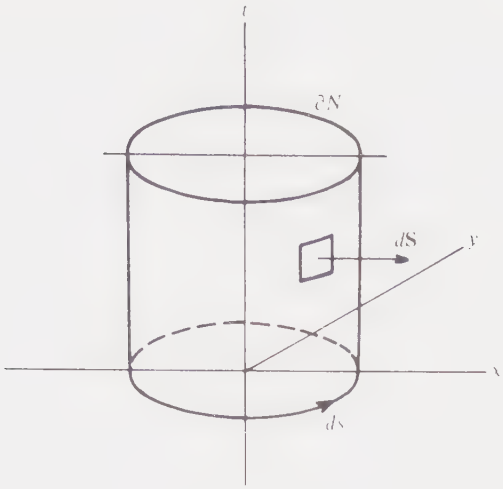



Figure 13 $N_3$ of (1.3.5; 2)

(the outer cylindrical surface plus top and bottom). (See Figure 13.) Then (1.3.4) is the integral form of the law of induction:

$$\int_{t_0}^{t_1} dt \int_{\substack{x^3=0 \\ |\mathbf{x}|=R}} d\mathbf{s}\cdot\mathbf{E} = \int_{\substack{t=t_0,\, x^3=0 \\ |\mathbf{x}|\le R}} d\mathbf{S}\cdot\mathbf{B} - \int_{\substack{t=t_1,\, x^3=0 \\ |\mathbf{x}|\le R}} d\mathbf{S}\cdot\mathbf{B}.$$

The vanishing of the differential of $F_{|N_3}$ means that

$$0 = d(F_{|N_3}) \Leftrightarrow \dot{B}_3 + (\nabla \times \mathbf{E})_3 = 0.$$

As already mentioned, closure is equivalent to exactness on starlike manifolds, and specifically on $\mathbb{R}^4$. This implies that there exists a

**Vector Potential** $A \in E_1(\mathbb{R}^4)$ (1.3.6)

$$F = dA. \tag{1.3.7}$$

In integrated form,

$$\int_{N_2} F = \int_{\partial N_2} A, \tag{1.3.8}$$

if $F \in E_2^0(N_2)$.

**Special Cases** (1.3.9)

1. $N_2 = \{(t, \mathbf{x}) \in \mathbb{R}^4 : t = x^3 = 0, |\mathbf{x}| \leq R\}$ (a disk)
   $\partial N_2 = \{(t, \mathbf{x}) \in \mathbb{R}^4 : t = x^3 = 0, |\mathbf{x}| = R\}$ (a circle).

   Then (1.3.8) expresses the magnetic flux as a line integral over the spatial part of the vector potential. Writing $A = A\,dt - \mathscr{A}\,d\mathbf{x}$,

$$\int_{\substack{|\mathbf{x}| \leq R \\ x^3 = t = 0}} dx^1 \wedge dx^2\, B_3 = \int_{\substack{|\mathbf{x}| = R \\ x^3 = t = 0}} d\mathbf{s} \cdot \mathscr{A}.$$

2. $N_2 = \{(t, \mathbf{x}) \in \mathbb{R}^4 : t_0 \leq t \leq t_1, x^1 = x^2 = 0\}$,
   $\partial N = \{(t, \mathbf{x}) \in \mathbb{R}^4 : t = t_0, x^1 = x^2 = 0\} \cup \{(t, \mathbf{x}) \in \mathbb{R}^4 : t = t_1,$
   $x^1 = x^2 = 0\}$.

   We have been looking only at compact submanifolds, thereby trivially fulfilling the requirement of compact support. If $F$ vanishes outside some compact region, or at least goes to zero fast enough at infinity, then (1.3.8) still holds, and implies

$$\int_{t_1}^{t_2} dt \wedge dx^3\, E_3 = \int_{t=t_0} dx^3\, \mathscr{A}_3 - \int_{t=t_1} dx^3\, \mathscr{A}_3.$$

**Remarks** (1.3.10)

1. Equation (1.3.7) determines $A$ only up to a gauge transformation $A \to A + d\Lambda$, where $\Lambda \in E_0(\mathbb{R}^4)$. Because $\partial(\partial N_2) = \varnothing$, there is no contribution from $\Lambda$ to the integral (1.3.8).
2. The statement (1.3.8) is on the whole stronger than (1.3.4). There exist manifolds $N_2$, without boundaries, that are not themselves boundaries of compact submanifolds $N_3$. For instance, the field of a magnetic monopole,

$$\mathbf{E} = \mathbf{0}, \qquad \mathbf{B} = \frac{e'}{4\pi} \frac{\mathbf{x}}{|\mathbf{x}|^3},$$

   satisfies (1.3.3) on $M = \{(t, \mathbf{x}) \in \mathbb{R}^4 : x \neq 0\}$, but gives rise to no vector potential: If $N_2 = \{(t, \mathbf{x}) \in \mathbb{R}^4 : t = 0, |\mathbf{x}| = R\}$, which is compact and has no boundary, then (1.3.2) would lead to the contradiction

$$0 = \int_{N_2} F = -\int_{|\mathbf{x}| = R} \mathbf{B} \cdot d\mathbf{S} = -e'.$$

   $N_2$ is in fact the boundary of $N_3 = \{(t, \mathbf{x}) \in \mathbb{R}^4 : t = 0, 0 < |\mathbf{x}| < R\}$, but since $N_3$ is not compact, one cannot conclude that $0 = \int_{N_3} dF = \int_{N_2} F$: for (1.3.4) to be applicable, $F$ must have compact support on $N_3$. The

physical significance of this is that any magnetic sources have been removed from $M$, whereas the field in $M$ is that of a magnetic monopole. There are theories that also claim there are no electric charges either, but that some complicated structure of $M$ makes it appear that there are electrical sources.

3. $M = \mathbb{R}^4 \setminus \{(t, \mathbf{x}) \in \mathbb{R}^4 : x^1 = x^2 = 0,\ x^3 \leq 0\}$ is starlike with respect to every point of the positive $z$-axis, and is contained in the $M$ of Remark 2. Therefore the field of a magnetic monopole satisfies $dF = 0$ on $M$, and there ought to be a vector potential. In fact,

$$\mathscr{A} = \frac{e'}{4\pi}\left(1 - \frac{z}{\sqrt{x^2 + y^2 + z^2}}\right)\frac{x\,dy - y\,dx}{x^2 + y^2}$$

reproduces the field of Remark 2 and is differentiable on $M$ (Problem 6). Equation (1.3.8) does not lead to a contradiction if $N_2 = \{(t, \mathbf{x}) \in \mathbb{R}^4 : t = 0,\ |\mathbf{x}| = R\} \cap M$, since $N_2$ is not compact. Physically, $A$ is the potential of an infinitely thin solenoid along the negative $z$-axis, which has been removed from $M$. We conclude that the empirical absence of magnetic monopoles indicates that the manifold in which we live has none of the pathologies described here, at least on the small scale.

The inhomogeneous equation describes how $F$ is produced by the sources, which are the charges or, more correctly, the charge densities. Since the charge density ought to yield the charge present in some three-dimensional, region when it is integrated over that region, we describe it as a 3-form $*J$. The corresponding 1-form $J$ consists of a temporal part, the charge density $\rho$, and a spatial part, the current density $\mathbf{J}$: $J = -\rho\,dt + \mathbf{J}\,d\mathbf{x}$. In terms of $J$ we may write down the

**Inhomogeneous Maxwell Equations** (1.3.11)

$$\delta F = J, \quad \text{or} \quad d{*}F = -{*}J; \tag{1.3.12}$$

and in integrated form,

$$-\int_{\partial N_3} {*}F = \int_{N_3} {*}J, \qquad F \in E_2^0(N_3). \tag{1.3.13}$$

**Remarks** (1.3.14)

1. The formulation of these equations requires a manifold with some additional structure, which defines the *-mapping. We do not require the concept of a covariant derivative, which will come up later.
2. Since none of these equations make reference to a chart, we are not restricted to the use of Cartesian coordinates. We shall even postulate the same form of Maxwell's equations for use on more general manifolds

than $\mathbb{R}^4$. This will cause the interaction with the gravitational field, and will make the equations describe such things as the bending of light by the sun.

3. The Poincaré transformations of $\mathbb{R}^4$ are the ones that leave the form of the metric $\eta_{ik}$, and consequently also * and Maxwell's equations, invariant.

The equations of (1.3.11) do not have solutions for an arbitrary 1-form $J$. On the contrary, because of (1.2.6(c)), their first corollary is

**Conservation of Charge** (1.3.15)

$$\delta J = 0, \quad \text{or} \quad d^*J = 0. \tag{1.3.16}$$

When integrated, this is

$$\int_{\partial N_4} {}^*J = 0, \qquad \text{if } {}^*J \in E_3^0(N_4). \tag{1.3.17}$$

**Special Cases** (1.3.18)

1. $N_4 = \{(t, \mathbf{x}): t_0 \le t \le t_1, |\mathbf{x}| \le R\}$. Then (1.3.17) implies that the charge that flows out through the surface equals the change of the charge contained in the sphere between the time $t_0$ and $t_1$:

$$\int\limits_{\substack{|\mathbf{x}|<R \\ t=t_0}} d^3x\, \rho - \int\limits_{\substack{|\mathbf{x}|<R \\ t=t_1}} d^3x\, \rho = \int_{t_0}^{t_1} dt \int\limits_{|\mathbf{x}|=R} d\mathbf{S}\cdot\mathbf{J}.$$

2. Let $\partial N = \{(t, \mathbf{x}): t = 0\} \cup \{(t, \mathbf{x}): \bar{t} \equiv (t - vx^1)/\sqrt{1 - v^2} = 0\} \equiv A \cup B$. This fails to be the boundary of a submanifold, because it has sharp bends (see Figure 14), though, as already mentioned, the bends could be smoothed out without changing the value of the integral much. It also fails to be compact, so let us assume that $J$ has spatially compact support. Then (recalling that ${}^*dx^0 = -dx^1 \wedge dx^2 \wedge dx^3$, so that $\int_{t=\text{const.}} {}^*J = \int d^3x\, J^0 = \int d^3x\, \rho$),

$$\int_A {}^*J = \int_B {}^*J \equiv Q.$$

If we calculate the integral over $B$ in Lorentz-transformed coordinates,

$$\bar{x}^1 = \frac{x^1 - vt}{\sqrt{1 - v^2}}, \qquad \bar{x}^2 = x^2, \qquad \bar{x}^3 = x^3, \qquad \bar{t} = \frac{t - vx^1}{\sqrt{1 - v^2}},$$

then we find that

$$Q \equiv \int_{t=0} d^3x\, J^0 = \int_{\bar{t}=0} d^3\bar{x}\, \bar{J}^0, \qquad \bar{J}^0 = \frac{J^0 - vJ^1}{\sqrt{1 - v^2}},$$

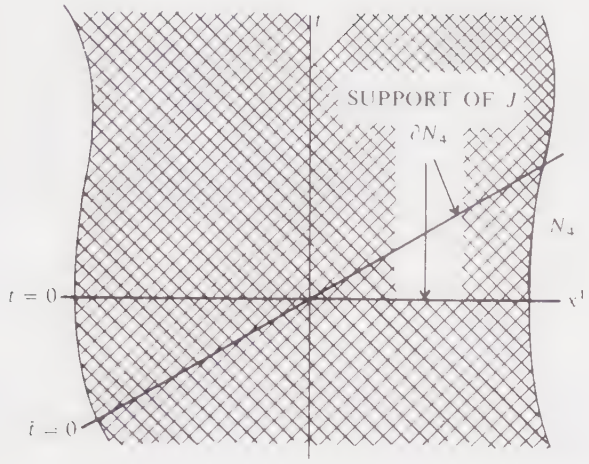



Figure 14 The Lorentz invariance of the total charge

since $(J^0, J^1)$, the components of $J$, transform as $(x^0, x^1)$. The total charge $Q$ is therefore a Lorentz invariant; the increase in the charge density by $\sqrt{1 - v^2}$ is compensated for by the Lorentz contraction of the volume element. One must bear in mind, however, that such concepts as scalar, vector, etc., are not defined for integrated quantities, because the transformation coefficients in general depend on the position.

**Remarks** (1.3.19)

1. The equations (1.3.11) imply more than (1.3.16); conservation of charge necessarily implies the existence of a 2-form $*F$ such that $*J = -d*F$ only if the manifold is starlike, or has some similar property (cf. (1.2.39; 12)). If $t$ is defined globally on $M$, and the submanifold $N_3 = \{t = 0\}$ is compact and has no boundary, then it follows from (1.3.11) that

$$Q = \int_{N_3} *J = \int_{\partial N_3} *F = 0,$$

since $\partial N_3 = \varnothing$. Hence a closed universe must have zero total charge, a conclusion that goes beyond mere charge conservation. The underlying reason is that in this case the lines of force emanating from a charge can not go to infinity, but must terminate in some opposite charge. The fact is that matter appears to be neutral to a remarkable degree. If the charge on the proton were increased relative to the electron charge, then the $1/r^2$ gravitational force would in effect be weakened for protons and strengthened for electrons. Since it is empirically known that the two forces are proportional to the masses and are weaker than the Coulomb force by a factor $10^{-36}$, matter is neutral to at least this accuracy.

2. If $F$ is a periodic 2-form and $N_3$ the periodicity volume then the contributions from the opposite boundaries to $\int_{\partial N_3} {}^*F$ cancel and $Q = 0$. We shall meet periodic $F$'s in plane wave solutions and thus the total charge in the periodicity volume has to vanish. The usual plane wave solutions actually solve Maxwell's equations for $J = 0$. But there are many wave-type solutions of a plasma coupled to the electromagnetic field and we see that one can get strict periodicity in this case only if particles of both signs of charge are present.

The physical manifestation of the field tensor $F$ is the electromagnetic force, i.e., the exchange of energy and momentum between the field and the charge-carriers. In field theory even the energy and momentum are introduced as local observables, and described by vector fields $\mathcal{T}^\alpha$, where $\mathcal{T}^0$ is the energy-current and $\mathcal{T}^j$, $j = 1, 2, 3$, are the momentum-currents. The components in a basis define the energy-momentum tensor $T^\alpha{}_\beta$:

$$\mathcal{T}^\alpha = T^\alpha{}_\beta e^\beta. \tag{1.3.20}$$

If the corresponding dual 3-forms are integrated over a spacelike† submanifold $N_3$, the result is the energy $P^0$ and the momentum $P^j$ contained in $N_3$:

$$P^\alpha = \int_{N_3} {}^*\mathcal{T}^\alpha. \tag{1.3.21}$$

As in classical mechanics, the Lagrangian formulation of the field equations makes it possible to construct a field energy.

In fact, there will be some close analogy to the expressions we met in Volume 1 except that field theory is richer in components. The coordinate $q$ is replaced by the vector potential $A$ and $\dot{q}$ by $dA = F$. The Lagrangian $\frac{1}{2}\dot{q}^2$ becomes the 4-form $-\frac{1}{2}F \wedge {}^*F$ and the dual $-{}^*F = \partial\mathcal{L}/\partial\, dA$ plays the role of $p = \partial L/\partial\dot{q}$. The sign is dictated by the requirement that the energy density $T_{00}$ be positive. The counterpart to the energy $H = \dot{q}p - L$ is

$$\begin{aligned} {}^*\mathcal{T}^\alpha &= -dA \wedge i^\alpha {}^*F + i^\alpha \mathcal{L}, \\ \mathcal{T}^\alpha &= \tfrac{1}{2}{}^*(i^\alpha F) \wedge {}^*F - (i^\alpha {}^*F) \wedge F \\ &= -(F^\alpha{}_\tau F^\tau{}_\beta + \tfrac{1}{4}\delta^\alpha{}_\beta F_{\rho\sigma}F^{\rho\sigma})\, dx^\beta, \end{aligned} \tag{1.3.22}$$

where $i^\alpha$ stands for the interior product (1.2.15) with $dx^\alpha$: $i^\alpha F = F^\alpha{}_\beta\, dx^\beta$. It has the job to pick out the components. We shall derive (1.3.22) in two ways. First by a consideration similar to classical mechanics where the invariance of the Lagrangian under translations produces a conservation law. Later, we shall see that it also equals $\partial\mathcal{L}/\partial e^\alpha$. This shows that change $e^\alpha \to L^\alpha{}_\beta e^\beta$ of the basis produces the homogeneous transformation law $\mathcal{T}^\alpha \to L^\alpha{}_\beta \mathcal{T}^\beta$ for the energy momentum currents of matter and of the

† A submanifold is called **spacelike (timelike, lightlike)** if $g$ is positive (negative, zero) when restricted to it.

electro-magnetic field. Only the Lagrangian of gravity contains $de$ which produces an inhomogeneous transformation.

As with charge-conservation, the change in the energy-momentum is determined by the codifferential of the currents, $\delta\mathscr{T}^\alpha$. The rate of change in time of $P^\alpha$ equals a four-dimensional integral over $d^*\mathscr{T}^\alpha$ and a surface integral over the spatial parts of the energy-momentum currents. For example, if $N_3 = \{(t, \mathbf{x}): t = \pm T, |\mathbf{x}| \leq R\}$, then

$$P^\alpha(T) - P^\alpha(-T) = \int\limits_{\substack{-T \leq t \leq T \\ |\mathbf{x}| \leq R}} d^*\mathscr{T}^\alpha - \int\limits_{\substack{-T \leq t \leq T \\ |\mathbf{x}| = R}} {}^*\mathscr{T}^\alpha. \tag{1.3.23}$$

If $F$ is coupled to a current, then energy and momentum are exchanged between the field and the current in amounts determined by the codifferential of the energy-momentum currents of the field. This codifferential is determined by Maxwell's equations and is known as the

**Lorentz Force** (1.3.24)

$$\delta\mathscr{T}^\alpha = \langle i^\alpha F | J \rangle$$

or, in components,

$$T^{\alpha\beta}{}_{,\beta} = F^{\alpha\beta} J_\beta.$$

**Derivation of the Equation for the Lorentz Force**

In the natural basis of $\mathbb{R}^n$,

$$L_{e^k}\omega \equiv (i^k \circ d + d \circ i^k)\omega_{j_1 \cdots j_p} e^{j_1 \cdots j_p} = \omega_{j_1 \cdots j_p}{}^{,k} e^{j_1 \cdots j_p},$$

and hence $L_{e^k}{}^*\omega = {}^*L_{e^k}\omega$. Defining $L^\alpha \equiv L_{e^\alpha}$, this means that

$$\begin{aligned} L^\alpha(F \wedge {}^*F) &= (L^\alpha F) \wedge {}^*F + F \wedge L^\alpha {}^*F \\ &= (L^\alpha F) \wedge {}^*F + F \wedge {}^*L^\alpha F = 2(L^\alpha F) \wedge {}^*F \end{aligned}$$

(cf. (1.2.18)(a) and (1.2.23)). Therefore, using (1.2.6) and (1.3.22),

$$\begin{aligned} d^*\mathscr{T}^\alpha &= -d[\tfrac{1}{2} i^\alpha (F \wedge {}^*F) - (i^\alpha F) \wedge {}^*F] \\ &= -\tfrac{1}{2} L^\alpha (F \wedge {}^*F) + (L^\alpha F) \wedge {}^*F - i^\alpha F \wedge d^*F = i^\alpha F \wedge {}^*J. \end{aligned}$$

According to (1.2.18) (b) and (c), $v \wedge {}^*\omega = {}^*i_v \omega (-1)^{p+1}$, for $v \in E_1$ and $\omega \in E_p$. Hence

$$\delta\mathscr{T}^\alpha = -{}^*d^*\mathscr{T}^\alpha = -{}^*(i^\alpha F \wedge {}^*J) = \langle i^\alpha F | J \rangle$$

**Remarks** (1.3.25)

1. It is important that the $i^\alpha$ in (1.3.22) be an interior product with a natural basis $e^\alpha$. The *-mapping does not commute with $L^\alpha$ in an arbitrary basis,

which spoils the above proof. It is also not possible for (1.3.24) to hold in every basis, because if $e^\alpha \to A^\alpha{}_\beta(x)e^\beta$, then the right side would transform linearly with $A$, while the left side would gain a term containing $dA$. A physical consequence of this is that fictitious forces must be added to the Lorentz force in an accelerated reference system.

2. If $z(s)$ is the world-line of a particle (see (I: §4.1)), then its current is

$$J_\beta(x) = e\int_{-\infty}^{\infty} ds\, \delta^4(x - z(s))\dot{z}_\beta(s),$$

where $\delta^4(x) \equiv \delta(x^0)\delta(x^1)\delta(x^2)\delta(x^3)$ (cf. Problem 7). The Lorentz force with this $J$,

$$e\int_{-\infty}^{\infty} ds\, \delta(x - z(s))F^{\alpha\beta}(z(s))\dot{z}_\beta(s),$$

equals $-\delta t^\alpha$, if we choose the energy-momentum currents of a particle so that the energy becomes $\int {}^*t^0 = m\dot{z}^0 > 0$:

$$\begin{aligned} t^\alpha &= +m\int_{-\infty}^{\infty} ds\, \dot{z}^\alpha\dot{z}_\beta(s)\delta^4(x - z)dx^\beta: \\ \delta t^\alpha &= +m\int_{-\infty}^{\infty} ds\, \dot{z}^\alpha\dot{z}^\beta\frac{\partial}{\partial x^\beta}\delta^4(x - z(s)) \\ &= -m\int_{-\infty}^{\infty} ds\, \dot{z}^\alpha\frac{d}{ds}\delta^4(x - z(s)) = m\int_{-\infty}^{\infty} ds\, \ddot{z}^\alpha\delta^4(x - z(s)) \\ &= e\int_{-\infty}^{\infty} ds\, \dot{z}_\beta F^{\beta\alpha}(z(s))\delta^4(x - z(s)). \end{aligned}$$

Thus the total energy and momentum are formally conserved. Yet the singularity of the field of a point-particle at $z(s)$ results in inconsistencies (cf. §2.4), which are resolved only for continuous matter in §3.1. Note that the Lorentz force is concentrated on the world-line of the particle, and that the equation $\delta(\mathcal{T}^\alpha + t^\alpha) = 0$, with the replacement of $\delta^4(x - z(s))$ by $\rho(x - z(s))$ for some continuous function $\rho$, is no longer true. Hence there is no local energy-momentum conservation for an extended charged particle, unless other forces hold it together.

If $\mathcal{T}^\alpha$ is defined in an arbitrary basis by (1.3.22), then it transforms as $e^\alpha$ under a change of basis:

$$e^\alpha \to \bar{e}^\alpha = A^\alpha{}_\beta e^\beta, \qquad \overline{\mathcal{T}}^\alpha = A^\alpha{}_\beta\mathcal{T}^\beta. \tag{1.3.26}$$

The energy and momentum are combined into a vector-valued 1-form. Thus, if $\delta(\mathcal{T}^\alpha + t^\alpha) = 0$, then a global Lorentz transformation treats $P^\alpha$ as a vector (Problem 5). If the $A^\alpha{}_\beta$ are allowed to depend on $x$, then this statement becomes meaningless, however, and the conservation equation $\delta(\overline{\mathcal{T}}^\alpha + \bar{t}^\alpha) = 0$ is false. In fact, in this case

$$d^*(\overline{\mathcal{T}}^\alpha + \bar{t}^\alpha) = dA^\alpha{}_\beta \wedge A^{-1\beta}{}_\gamma{}^*(\overline{\mathcal{T}}^\gamma + \bar{t}^\gamma).$$

Choosing $e^\alpha$ as the Cartesian $dx^\alpha$ of $\mathbb{R}^4$, i.e., setting the $\omega^\alpha{}_\beta$ of (1.2.25) to zero, makes

$$\bar{\omega}^\alpha{}_\gamma = -(dA^\alpha{}_\beta)(A^{-1})^\beta{}_\gamma, \tag{1.3.27}$$

according to (1.2.26; 3), and therefore

$$\delta(\bar{\mathcal{T}}^\alpha + \bar{t}^\alpha) = -\langle \bar{\omega}^\alpha{}_\gamma | \bar{\mathcal{T}}^\gamma + \bar{t}^\gamma \rangle \tag{1.3.28}$$

**Remarks** (1.3.29)

1. It seems odd at first that with general bases there are only nonconservation theorems; they are due to fictitious forces in accelerated reference frames and hence are to be expected on physical grounds. They provide a clue to an understanding of the structure of Einstein's equations.
2. If (1.3.28) is compared with (1.3.24), one sees that $\omega$ plays the same role for fictitious forces as $F$ plays for the Lorentz force. $\omega$ acts on $\mathcal{T} + t$ as $F$ acts on $J$.

**Example** (1.3.30)

A rotating basis. Let

$$\begin{aligned} \bar{e}^0 &= dt \\ \bar{e}^1 &= dx \cos \nu t + dy \sin \nu t \\ \bar{e}^2 &= -dx \sin \nu t + dy \cos \nu t \\ \bar{e}^3 &= dz. \end{aligned}$$

This basis is orthogonal but not natural;

$$d\bar{e}^1 = \nu\, dt \wedge \bar{e}^2, \qquad d\bar{e}^2 = -\nu\, dt \wedge \bar{e}^1 : \bar{\omega}^1{}_2 = -\nu\, dt = -\bar{\omega}^2{}_1.$$

It makes

$$\begin{aligned} d(*\mathcal{T}_1 + *t_1) &= \nu\, dt \wedge (*\mathcal{T}_2 + *t_2) \\ d(*\mathcal{T}_2 + *t_2) &= -\nu\, dt \wedge (*\mathcal{T}_1 + *t_1); \end{aligned}$$

or, if $P_n(t) \equiv \int_{x^0=t} (*\mathcal{T}_n + *t_n)$, then it follows from Stokes's theorem that

$$P_1(t_1) - P_1(t_0) = \nu \int_{t_0}^{t_1} dt\, P_2(t)$$

$$P_2(t_1) - P_2(t_0) = -\nu \int_{t_0}^{t_1} dt\, P_1(t).$$

These are just like the equations $\dot{P}_1 = \nu P_2$, $\dot{P}_2 = -\nu P_1$ of the mechanics of point particles (cf. (I: 3.2.15; 2)) in a rotating system.

Einstein's theory introduces gravity in (1.3.28) via the principle of equivalence. The gravitational potential is represented by the metric $g$. In other

words, if we use orthogonal bases†, so that $g = e^\alpha \otimes e^\beta \eta_{\alpha\beta}$, then the 1-forms $e^\alpha$ are analogous to the $A$ of (1.3.7). Formula (1.3.28) replaces the Lorentz force (1.3.24), and is assumed to hold not only in Minkowski space, but also in the pseudo-Riemannian space determined by $g$.

**Remarks** (1.3.31)

1. As there are four 1-forms $e^\alpha$ representing gravitational potentials, there is not just one current, but four entering into the force.
2. The $\omega$'s are the new field strength, although they do not transform linearly under a change of basis, as $F$ does, but inhomogeneously according to (1.2.26; 3). As a result, they can always be transformed to zero at a point. A physical realization of such a system would be a freely falling elevator, in which there is no gravitational force.
3. Theories have recently been proposed [16] in which the $\omega$'s are analogous to $A$; however, it would take us too far afield to explore this analogy.

If $\omega$ plays the role of $F$, then

$$de^\alpha = -\omega^\alpha{}_\beta \wedge e^\beta, \qquad \omega_{\alpha\beta} = -\omega_{\beta\alpha},$$

can be viewed as the counterpart of the relation $F = dA$. To construct the inhomogeneous equations, one might try to equate the codifferentials of 2-forms linear in $\omega$ to the energy and momentum currents.

Since now there are four currents they ought to be codifferentials of the same number of 2-forms representing the field strength. They should be linear in the $\omega$'s but cannot be identical because the $\omega$'s are six independent 1-forms. The right combination turns out to be

$$F^\gamma = -\tfrac{1}{2} i_{\omega_{\alpha\beta}} e^{\alpha\beta\gamma}. \tag{1.3.32}$$

Einstein's equations are the counterpart to inhomogeneous Maxwell's equations and assert that $\delta F^\gamma$ equals the energy momentum currents. However, there is the problem that according to (1.3.28) the currents of matter and electromagnetism alone are not conserved in an arbitrary basis and $\delta F^\gamma = \text{const}(\mathcal{T}^\gamma + t^\gamma)$ would be mathematically inconsistent. This non-conservation is due to fictitious gravitational forces and can be compensated by adding to the contributions from matter and electromagnetism the energy momentum currents of gravity $\ell^\alpha$. It turns out that they are constructed following the pattern of (1.3.22) where $de^\alpha$ replaces $dA$ and $F^\alpha$ replaces $F$:

$${}^*\ell^\alpha = \frac{1}{16\pi\kappa}[de^\beta \wedge i^\alpha F_\alpha - i^\alpha \mathcal{L}_g].$$

† As will be done throughout the rest of this chapter. Then indices are raised and lowered by means of $\eta$: $\omega_{\alpha\beta} = \eta_{\alpha\gamma}\omega^\gamma{}_\beta$, etc., which at most changes some signs.

Here $\kappa$ is the gravitational constant which relates the geometric to the physical quantities. The gravitational Lagrangian

$$\mathcal{L}_g = \frac{1}{16\pi\kappa} {}^*e^{\beta\gamma} \wedge \omega_\gamma{}^\alpha \wedge \omega_{\alpha\beta}$$

is such that

$${}^*F^\alpha = \frac{\partial \mathcal{L}_g}{\partial de_\alpha}.$$

These calculations will be done in detail in §4.2. For the moment we just want to show in analogy to the inhomogeneous Maxwell equation we have

**Einstein's Equations** (1.3.33)

$$\partial F^\gamma = \tfrac{1}{2}\delta(i_{\omega_{\alpha\beta}} e^{\alpha\beta\gamma}) = 8\pi\kappa(\mathcal{T}^\gamma + t^\gamma + \ell^\gamma).$$

**Remarks** (1.3.34)

1. The analogy with (1.3.12) is not quite perfect. Einstein's equations do not contain the codifferentials of $de^\alpha = -{}^*i_{\omega\gamma_\beta}{}^*e^\beta$, but, instead, the 3-form ${}^*e^\beta$ is replaced with $e^{\alpha\beta\gamma}$. The left side of (1.3.33) equals $\Delta e^\gamma$ only in special bases.
2. In §4.2 (1.3.33) is compared with the form used by Einstein where only the curvature appears.
3. Equation (1.3.28) follows from (1.3.33) in the same way as the Lorentz force follows from Maxwell's equations.
4. Since $\ell$ contains $\omega$, it does not transform under a change of basis according to the linear law (1.3.26). Such a law would be inconsistent with the continuity equation

$$\delta(\mathcal{T}^\alpha + t^\alpha + \ell^\alpha) = 0$$

arising from (1.3.33). In particular, in a basis where all the $\omega$'s vanish at some point, $\ell^\alpha$ also vanishes at that point. Physically speaking, this is a consequence of the principle of equivalence: Since gravity can be transformed away at any point, it can have neither energy nor momentum definitely localized at a point. On the other hand, $\ell^\alpha$ may be nonzero even in the absence of gravity (flat space), if an unnatural basis is used. This is to say that fictitious forces produce fictitious energies and momenta.
5. Under a change of basis, Equations (1.3.33) transform linearly because the inhomogeneous contribution to $\ell^\gamma$ is compensated for in the transformation of the left sides of the equations.

6. There is, of course, not just one $\ell^\gamma$ for which the continuity equation (1.3.28) holds; it is always possible to add a closed form to it. With an orthogonal basis, our $\ell^\gamma$ agrees with the pseudotensor used by Landau and Lifshitz. At this stage, our definition of $\ell^\gamma$ as the source of the term $i_{\omega_{\alpha\beta}} e^{\alpha\beta\gamma}$ may not seem convincing; we shall see later that in some familiar cases it reproduces the gravitational energy of the elementary theory, in a basis that is as Cartesian as possible.

While nonlinearity makes the equations harder to solve, there is no difficulty in drawing general conclusions by using Stokes's theorem and restricting to submanifolds. As with the inhomogeneous Maxwell equations, we obtain some

**Corollaries** (1.3.35)

1. $$\int_{\partial N_4} (*\mathcal{T}^\gamma + *t^\gamma + *\ell^\gamma) = 0.$$

If a timelike coordinate $t$ is specified globally on $M$, and the energy and momentum fall off sufficiently fast at infinity on the submanifold $t =$ constant, then the total energy and momentum are conserved:

$$P^\gamma(t_1) = P^\gamma(t_2),$$

$$P^\gamma(t_1) = \int_{t=t_1} (*\mathcal{T}^\gamma + *t^\gamma + *\ell^\gamma).$$

2. Equations (1.3.33) allow us to make an even stronger statement,

$$\int_{N_3} (*\mathcal{T}^\gamma + *t^\gamma + *\ell^\gamma) = -\frac{1}{16\pi\kappa} \int_{\partial N_3} \omega_{\alpha\beta} \wedge *e^{\alpha\beta\gamma}.$$

More specifically, if the submanifold $t = t_0$ is compact, space-like, and has no boundary, then

$$P^\gamma(t_0) = 0.$$

In analogy with (1.3.19; 1), this means that the total energy and momenta of a closed universe are zero.

**Remarks** (1.3.36)

1. A rough, order-of-magnitude estimate shows that the negative gravitational energy may well balance the rest energy of the matter in the universe. The average distance between galaxies is about $6 \times 10^6$ light-years. Since the universe is some $10^{10}$ years old and about as many light-years across, there are $\sim 10^{11}$ galaxies in all. If each one has about

$10^{11}$ stars at $10^{33}$ grams apiece, then the universe has a mass of perhaps $10^{55}$ grams. Hence the gravitational energy (in cgs units) is on the order of

$$-M\frac{M\kappa}{R} \sim -M \cdot 10^{55} \cdot 10^{-7} \cdot 10^{-28} \sim -M \cdot 10^{20},$$

where the radius of the universe $R \sim 10^{10}$ light-years $\sim 10^{28}$ cm. This falls short of the rest-energy $Mc^2 \sim M \times 10^{21}$ by about a factor of 10, but there are large uncertainties in our calculation, owing to our ignorance about how much of the energy of the universe is detectable. Later we shall again have occasion to convince ourselves that $\ell^\gamma$ indeed contributes a gravitational energy $\sim -M^2\kappa/R$.

2. Because of the complicated dependence on the basis, there is no intuitively appealing observable corresponding to $\ell^\gamma$. The total energy and momentum $P^\gamma$ are less sensitive to the choice of basis, as they may be expressed as integrals over $\partial N_3$ according to (1.3.35; 2). Changes of basis leave them invariant as long as nothing is changed on the boundary. If the space is asymptotically a Minkowski space, then these observables transform as vectors under any transformation that asymptotically approaches a Lorentz transformation.
3. The question of when the energy-momentum forms can be defined globally even if it requires more than one chart to describe the manifold will be discussed later (4.2.12; 8).

So much for the structure of Maxwell's and Einstein's equations. In the remainder of the book, after this preview, we study them in greater detail, and will not only guess a few particular solutions, but also see what general statements can be made about the solutions of these equations.

**Problems** (1.3.37)

1. What are the consequences of (1.3.13) for $N_3 = \{(t, \mathbf{x}): t = 0, |\mathbf{x}| \leq R\}$, and of (1.3.12) restricted to $N_3$?

2. The same question for $N_3 = \{(t, \mathbf{x}): t_0 \leq t \leq t_1, x^3 = 0, |\mathbf{x}| \leq R\}$.

3. According to (1.3.18; 3), there is a Lorentz-transformed system with a nonzero charge-density $\bar{J}_0$, if $J_0 = 0$ but $J_1 \neq 0$. Explain physically how a neutral system with a current appears charged when seen from a moving reference frame.

4. Show that on $\mathbb{R}^n$ the Lie derivative with respect to the natural basis, $L_k = i_{dx_k} \cdot d + d \circ i_{dx_k}$, of $e^{j_1 \cdots j_p}$ simply yields

$$L_k \omega_{j_1 \cdots j_p} e^{j_1 \cdots j_p} = \omega_{j_1 \cdots j_p, k} e^{j_1 \cdots j_p},$$

in the natural basis.

5. Discuss how $P^\alpha \equiv \int_{N_3} {}^*\mathcal{T}^\alpha$ is affected by Lorentz transformations.

6. Calculate $F$ with the potential given in (1.3.10; 3).

7. Verify that $\delta J = 0$ for a current of the form

$$J_\alpha(x) = \int_{-\infty}^{\infty} ds\, \dot{z}_\alpha(s)\rho(x - z(s)), \qquad \rho \in E_0^0(\mathbb{R}^4).$$

**Solutions** (1.3.38)

1.
$$-\,{}^*F = \begin{pmatrix} 0 & B_1 & B_2 & B_3 \\ -B_1 & 0 & E_3 & -E_2 \\ -B_2 & -E_3 & 0 & E_1 \\ -B_3 & E_2 & -E_1 & 0 \end{pmatrix}$$

Thus (1.3.13) implies that

$$\int_{|\mathbf{x}|=R,\, t=0} d\mathbf{S}\cdot\mathbf{E} = \int_{|\mathbf{x}|\le R,\, t=0} d^3x\, \rho(x); \quad \text{and} \quad d({}^*F_{|N_3}) = -{}^*J_{|N_3}$$

means that $\nabla\cdot\mathbf{E} = \rho$.

2.
$$\int_{t_0}^{t_1} dt \int_{\substack{x^3=0\\ |\mathbf{x}|=R}} d\mathbf{s}\cdot\mathbf{B} = \int_{\substack{t=t_1,\, x^3=0\\ |\mathbf{x}|\le R}} d\mathbf{S}\cdot\mathbf{E} - \int_{\substack{t=t_0,\, x^3=0\\ |\mathbf{x}|\le R}} d\mathbf{S}\cdot\mathbf{E} + \int_{t_0}^{t_1} dt \int_{\substack{x^3=0\\ |\mathbf{x}|\le R}} d\mathbf{S}\cdot\mathbf{j},$$

or $\nabla\times\mathbf{B} = \mathbf{j} + \dot{\mathbf{E}}$.

3. A neutral current consists of oppositely charged currents flowing past each other. Because of the differing definitions of simultaneity, the two currents seem to have different charges in the moving frame (see Figure 15).

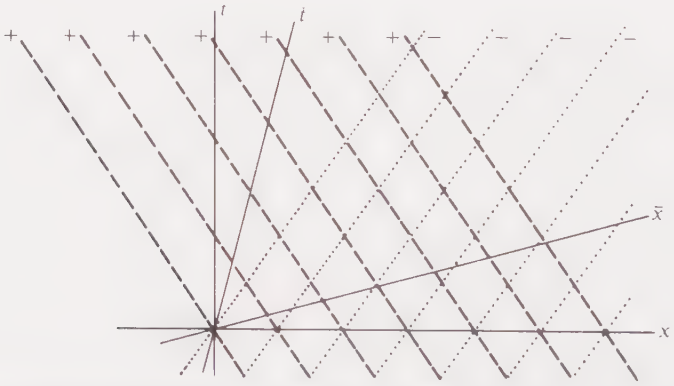



Figure 15 At any point of time $t$ = constant, $x$ sees 7 − 's to 7 + 's. At any point of time $\bar{t}$ = constant, $\bar{x}$ sees 5 − 's to 7 + 's.

4.
$$i^k\, d\omega = \omega_{j_1\cdots j_p,j} i^k e^{jj_1\cdots j_p} = \omega_{j_1\cdots j_p,}{}^k e^{j_1\cdots j_p} - p\omega^k{}_{j_2\cdots j_p,j} e^{jj_2\cdots j_p}$$
$$di^k\omega = pd(\omega^k{}_{j_2\cdots j_p} e^{j_2\cdots j_p}) = p\omega^k{}_{j_2\cdots j_p,j} e^{jj_2\cdots j_p}.$$

5. Since the natural basis transforms according to $d\bar{x}^\alpha = L^\alpha{}_\beta\, dx^\beta$, it is also true that $\bar{\mathscr{T}}^\alpha = L^\alpha{}_\beta \mathscr{T}^\beta$. Now note that $N_3$ looks different in the new coordinates. For instance, if $N_3$ is the hyperplane $t = 0$, then in the new system it is not $\bar{t} = 0$, but $\bar{t} = v\bar{x}$. It is only if $\delta\mathscr{T}^\alpha = 0$ and $\mathscr{T}^\alpha$ vanishes fast enough at infinity that

$$\int_{t=0} {}^*\mathscr{T}^\alpha = \int_{\bar{t}=0} {}^*\bar{\mathscr{T}}^\alpha$$

(cf. (1.3.18; 2)).

6. $r^2 \equiv x^2 + y^2 + z^2$: then

$$B_x = -\frac{\partial \mathscr{A}_y}{\partial z} = \frac{e^1}{4\pi}\frac{x}{r}\frac{1}{x^2+y^2}\left(1 - \frac{z^2}{r^2}\right) = \frac{e^1}{4\pi}\frac{x}{r^3}$$

$$B_y = \frac{\partial \mathscr{A}_x}{nz} = \frac{e^1}{4\pi}\frac{y}{r^3}$$

$$B_z = \frac{\partial \mathscr{A}_y}{\partial x} - \frac{\partial \mathscr{A}_x}{\partial y} = \frac{e^1}{4\pi}\left\{\frac{z(x^2+y^2)}{r^3(x^2+y^2)} + \frac{1}{x^2+y^2} - \frac{2x^2}{x^2+y^2} + \frac{1}{x^2+y^2} - \frac{2y^2}{x^2+y^2}\right\} = \frac{e^1}{4\pi}\frac{z}{r^3}.$$

7. $$\frac{\partial}{\partial x^\beta}\int_{-\infty}^{\infty} ds\, \rho(x - z(s))\dot{z}^\beta(s) = -\int_{-\infty}^{\infty} ds\, \dot{z}^\beta(s), \frac{\partial}{\partial z^\beta}\rho(x - z(s))$$

$$= -\int_{-\infty}^{\infty} ds \frac{\partial}{\partial s}\rho(x - z(s)) = 0.$$

No boundary terms appear from the partial integration, if the world-lines are infinite and $\rho \in E_0^0$. The normalization $\langle \dot{z} | \dot{z} \rangle = -1$ is irrelevant: $\delta J = 0$ even if the particles move faster than light.

# 2 The Electromagnetic Field of a Known Charge Distribution

## 2.1 The Stationary-Action Principle and Conservation Theorems

*If the field equations originate from a stationary-action principle, then a conserved current can be constructed for each parameter of an invariance group.*

Field theory may be regarded as a generalization of the mechanics of point particles, in which the dynamical variables $q_i(t)$ are replaced with fields $\Phi(x, t)$, such as $\mathbf{E}(x, t)$ and $\mathbf{B}(x, t)$. The discrete index $i$ goes over to the continuous variable $x$, and, accordingly, the sum $\sum_i$ is replaced with an integral $\int d^3x$. A direct transcription of the formalism of I, §3, leads to infinite-dimensional manifolds, which we would prefer to avoid. Instead, we merely generalize the stationary-action principle (I:2.3.20) in order to find the analogues of the constants arising from the invariance properties. It is clear that in field theory the action $\int dt\, L(q, \dot{q})$ involves an integral over a four-dimensional submanifold $N_4$, and thus requires a 4-form, which allows the construction of a chart-independent integral.

**The Lagrangian Formulation of Field Theory** (2.1.1)

The action is given by

$$W = \int_{N_4} \mathscr{L}(\Phi, d\Phi),$$

where $\mathscr{L} \in E_4$ is the **Lagrangian**. The field equations result from the requirement that $\delta W = 0\ \forall N_4$ compact and $\forall \delta\Phi$ such that $\delta\Phi_{|\partial N_4} = 0$.†

If we strengthen the homogeneous Maxwell equations to $F = dA$, then in pseudo-Riemannian space, the appropriate

**Electromagnetic Lagrangian** (2.1.2)

*is*

$$\mathscr{L} = -\tfrac{1}{2}\, dA \wedge {}^*dA - A \wedge {}^*J.$$

**Proof**

Making a variation $A \to A + \delta A$ and using (1.2.18) (a), one finds

$$-\delta W = \int_{N_4} \delta A \wedge [{}^*J + d{}^*dA] + \int_{\partial N_4} \delta A \wedge {}^*dA,$$

which vanishes if $\delta A_{|\partial N_4} = 0$ and $d{}^*F = -{}^*J$. □

**Remarks** (2.1.3)

1. The variational formulation offers no guarantee of existence or uniqueness of the solutions of the field equations. Nowhere has it been assumed that $d{}^*J = 0$, though without this condition it is not possible to satisfy $\delta W = 0\ \forall \delta A$ such that $\delta A_{|\partial N_4} = 0$. The reason is easy to discover. With the gauge transformation $A \to A + d\Lambda$, where $\Lambda_{|\partial N_4} = 0$, $W$ changes by $\int_{N_4} \Lambda\, d{}^*J$, and is linear in $\Lambda$ not only for infinitesimal $\Lambda$. As a linear functional, either $W$ has no stationary points, or else, if $d{}^*J = 0$, it has a plateau. Accordingly, either there are no solutions at all, or else the solution is not uniquely fixed by any boundary condition whatsoever, because there is always the possibility of a gauge transformation.
2. According to (I: 5.2.8),

$$-\tfrac{1}{2}F \wedge {}^*F = -\tfrac{1}{4}F_{\sigma\rho}F^{\sigma\rho}{}^*\mathbf{1} = \tfrac{1}{2}(|\mathbf{E}|^2 - |\mathbf{B}|^2){}^*\mathbf{1}.$$

The sign of $\mathscr{L}$ has been chosen so that the interaction

$$-A \wedge {}^*J = -{}^*i_J A = -J^\alpha A_\alpha {}^*\mathbf{1}$$

of a point particle moving along the world-line $z(s)$ (cf. (1.3.25; 2)) has the same sign

$$-e \int_{-\infty}^{\infty} ds\, \dot{z}^\alpha(s) A_\alpha(z(s)) \delta^4(x - z(s)) {}^*\mathbf{1}$$

† We use the symbol $\delta$ in § 2.1 for variations, rather than for codifferentials.

as in (I: 5.1.8). If a term $\frac{1}{2}m\int ds\, \dot{z}^{\alpha}(s)\dot{z}_{\alpha}(s)$ were added to the action, then both the field equations and the equation of motion could be derived from the same stationary-action principle by varying $A(x)$ and $z(s)$. The coupled system of equations suffers from difficulties due to the reaction of a particle on itself, as will be discussed more fully in §2.4.

The advantage of the Lagrangian formulation is that every one-parameter invariance group of $\mathscr{L}$ furnishes a conservation theorem. In field theory the argument goes as follows: If the 4-form $\mathscr{L}$ depends on the $p$-form $\Phi_j$, then a variation of $\mathscr{L}$ (taking $\partial\mathscr{L}/\partial\Phi_j$ with $\delta\Phi$ commuted through to the left) produces

$$\delta\mathscr{L} = \sum_j \delta\Phi_j \wedge \left[\frac{\partial\mathscr{L}}{\partial\Phi_j} - (-1)^p d\,\frac{\partial\mathscr{L}}{\partial(d\Phi_j)}\right] + d\left(\sum_j \delta\Phi_j \wedge \frac{\partial\mathscr{L}}{\partial(d\Phi_j)}\right), \quad (2.1.4)$$

and the field equations require that the term in the square brackets [ ] is zero. If the variation $\delta$ is that of a Lie derivative $L_v$, $v \in E_1$, then by (1.2.23), $\delta\mathscr{L} = L_v\mathscr{L} = di_v\mathscr{L}$, because the exterior differential of a 4-form vanishes on a four-dimensional manifold. (We think of $\delta$ as a kind of derivative, and not as an infinitesimal quantity.) Hence (2.1.4) says that a certain 3-form is closed and thus, by use of the * mapping, that there is a conserved current.

**Noether's Theorem** (2.1.5)

*Suppose that* $v \in E_1$, $\delta\Phi = L_v\Phi$ *and* $\delta\mathscr{L} = L_v\mathscr{L}$. *Then*

$$d\left[\sum_j L_v\Phi_j \wedge \frac{\partial\mathscr{L}}{\partial(d\Phi_j)} - i_v\mathscr{L}\right] \equiv -d^*\mathscr{T}_v = 0.$$

**Remarks** (2.1.6)

1. The validity of (2.1.5) is premised on the variation $\delta$ including everything that is affected by $L_v$. For instance, even if $J = 0$, the $\mathscr{L}$ of (2.1.2) involves not only $dA$ but also the metric $g$, through the * mapping, and $g$ was not allowed to vary below (2.1.2). Since we supposed that $\delta{*}F = {*}\delta F$, Noether's theorem (2.1.5) applies to (2.1.2) only if $J = 0$ and $L_v{*} = {*}L_v$.
2. If $L_v$ is a translation, then (2.1.5) defines what is known as the **canonical energy-momentum tensor**.
3. The field equations remain unchanged when an exact 4-form is added to $\mathscr{L}$, $\mathscr{L} \to \mathscr{L} + dG$, where $G \in E_3$, since the addition only contributes a boundary integral to $\delta W$. However, the 3-form in the square brackets in (2.1.5) gets a contribution, which, if $G$ depends only on the fields $\Phi$ and not on their derivatives, is $di_v G$ (Problem 1). As an exact 3-form, it does not contribute to integrals over submanifolds without boundaries, but it can affect the conserved observables locally. This difficulty is not

encountered in classical mechanics, which is formally a one-dimensional field theory, and where $G$ would be in $E_0$, and hence $i_v G = 0$. Indeed, an additional $dG$ such that

$$\frac{d}{dt} G(q) = \dot{q}_i \frac{\partial G}{\partial q_i}$$

does not change the Hamiltonian

$$H = \dot{q}_i \frac{\partial \mathscr{L}}{\partial \dot{q}_i} - \mathscr{L}$$

at all.

**Application to the Electromagnetic Field** (2.1.7)

If $J = 0$ and $L_v{}^* = {}^*L_v$, then

$$\begin{aligned} &d[-(L_v A) \wedge {}^*dA + \tfrac{1}{2} i_v (dA \wedge {}^*dA)] \\ &\quad = d[-\tfrac{1}{2}(i_v F) \wedge {}^*F + \tfrac{1}{2} F \wedge i_v {}^*F - (di_v A) \wedge {}^*F] = 0. \end{aligned}$$

**Remarks** (2.1.8)

1. As can be seen above, if $r = e^x$, then in addition to the electromagnetic energy-momentum forms ${}^*\mathscr{T}^\alpha$ given in (1.3.22), there is a new term $(di_v A) \wedge {}^*F$. If $J = 0$, then this is an exact 3-form, $(di_v A) \wedge {}^*F = d(i_v A \wedge {}^*F) + i_v A \wedge {}^*J$, so that (2.1.7) is a special case of (1.3.24). Unfortunately, the new term contains $A$ as well as $F$, and thus depends on the gauge. This does not contradict the gauge-invariance of $\mathscr{L}$ (assuming $J = 0$): Even in the mechanics of point particles the angular momentum $m[\mathbf{x} \times \dot{\mathbf{x}}]$ fails to be translation-invariant, although $\mathscr{L} = m|\dot{\mathbf{x}}|^2/2$ is.
2. The nonuniqueness mentioned in (2.1.6; 3) consists of an additional $d(F \wedge A) = F \wedge F$ in $\mathscr{L}$. Since the corresponding $G$ depends on $dA$ as well as $A$, the extra term in the conserved observable is changed to $2d((i_v A) \wedge F)$ (Problem 5). This is conserved independently of whether ${}^*L_v = L_v{}^*$, because $F \wedge F$ makes no reference to the metric structure of space-time. The expression $d((i_v A) \wedge F)$ not only depends on the gauge, but it also has the wrong reflection property. For example, the energy density would have a term $\sim \mathbf{B} \cdot \nabla V$, which changes sign if $(t, \mathbf{x}) \to (t, -\mathbf{x})$, as $(V, \mathscr{A}) \to (V, -\mathscr{A})$ and $(\mathbf{E}, \mathbf{B}) \to (-\mathbf{E}, \mathbf{B})$. In the so-called gauge theories, similar expressions determine what are known as topological charges $\int F \wedge F$, which characterize the topological structure of the bundle.
3. The local energy is defined by the coupling with gravity, and the formula for it is derived by letting the metric vary in $\mathscr{L}$. This variation has no

contribution from $F \wedge F$, and $F \wedge {}^*F$ gives rise only to the $\mathscr{T}^z$ of (1.3.22), which is gauge-invariant (see (4.2.8)).

4. If $J \neq 0$, then adding $L_v {}^*J$ to the variations in (2.1.2) results in the Lorentz force. If $d{}^*F = -{}^*J$, then from (2.1.2),

$$\begin{aligned} L_v \mathscr{L} &= -d[(L_v A) \wedge {}^*dA] - A \wedge L_v {}^*J \\ &= -d[(i_v F) \wedge {}^*F] - (di_v A) \wedge {}^*J - A \wedge L_v {}^*J. \end{aligned}$$

On the other hand, due to the rules governing $L_v$,

$$L_v \mathscr{L} = -\tfrac{1}{2} di_v(F \wedge {}^*F) - L_v(A \wedge {}^*J).$$

Together, these facts imply the formula

$$-d{}^*\mathscr{T}_v \equiv \tfrac{1}{2} d[F \wedge i_v {}^*F - (i_v F) \wedge {}^*F] = -(i_v F) \wedge {}^*J$$

for any vector field such that ${}^*L_v = L_v {}^*$.

5. If $d{}^*J = 0$ and $\mathscr{L}$ is as in (2.1.2), then $\int_{N_4} \mathscr{L}$ is invariant under $A \to A + d\Lambda$, $\Lambda_{|\partial N_4} = 0$. The gauge group is then a huge (Abelian) invariance group, and is not even locally compact, since it contains arbitrary functions. This leads one to think that there must be an infinite number of conservation theorems. However, they always reduce to trivialities, or rather to identities that hold independently of the field equations. By the way, this is not a characteristic only of field theories, but also occurs in point-particle mechanics (Problem 3). In (2.1.2) the variation $\delta A = d\Lambda$ (and $d{}^*J = 0$) produces

$$0 = \delta W = -\int_{N_4} \Lambda d(d{}^*F) + \int_{\partial N_4} \Lambda({}^*J + d{}^*F).$$

Since $\Lambda_{|\partial N_4} = 0$, this means that $d(d{}^*F) = 0$, which is true regardless of whether $d{}^*F = -{}^*J$.

As we have seen, there is a conservation theorem for each $v$ whose Lie derivative does not destroy the structure of $\mathscr{L}$ determined by the metric. Such vector fields are important enough to merit a

**Definition** (2.1.9)

A vector field $v$ satisfying $L_v g = 0$ on a pseudo-Riemannian manifold with the metric $g$ is known as a **Killing vector field**.

**Remarks** (2.1.10)

1. Because $L_x L_y - L_y L_x = L_{[x,y]}$, where $[\ ,\ ]$ is the Lie bracket of vector fields (see (I: 2.5.9; 6)), the Killing vector fields form a Lie algebra with $[\ ,\ ]$. (But not a module: $fv$, with $f \in E_0$ is not necessarily a Killing vector field if $v$ is.)

2. If an orthogonal basis $e^i$ ($g = e^i \otimes e^j \eta_{ij}$) is used to decompose the Lie derivative of the $e^i$ as $L_v e^i = A^i{}_j e^j$, $A^i{}_j \in E_0$, then $v$ is a Killing vector field iff $A_{ij} = -A_{ji}$, where $A_{ij} \equiv \eta_{ik} A^k{}_j$.
3. Problem 2 is to show that $*L_v = L_v *$ for Killing vector fields $v$.
4. It is possible for $*L_v \omega$ to equal $L_v * \omega$, $\omega \in E_p$ (for particular values of $p$), even if $v$ is not a Killing vector field. For example, $L_v e^j = f e^j, f \in E_0$, makes $L_v e^{j_1 \cdots j_p} = p f e^{j_1 \cdots j_p}$ and $L_v * e^{j_1 \cdots j_p} = (m-p) f * e^{j_1 \cdots j_p}$, and so $*L_v \omega = L_v * \omega$ for all $\omega \in E_{m/2}$. Yet $L_v$ generates the conformal transformation $L_v g = 2fg$, and $v$ is not a Killing vector field.

**Examples** (2.1.11)

$M$ is taken as $\mathbb{R}^4$ with $e^\alpha = dx^\alpha$ and $g = e^\alpha \otimes e^\beta \eta_{\alpha\beta}$ in these examples.

1. The rigid displacement $v = e^\alpha$ leaves $g$ invariant:

$$L_v e^\gamma = d i_{e^\alpha} e^\gamma + i_{e^\alpha} de^\gamma = 0.$$

The $\mathscr{T}_v$ of (2.1.8; 4) becomes the $\mathscr{T}^\alpha$ of (1.3.22), and (2.1.8; 4) reduces to the Lorentz force (1.3.24).

2. $v = x^\beta e^\alpha - x^\alpha e^\beta$ generates a Lorentz transformation

$$L_v e^\gamma = x^\beta i_{e^\alpha} de^\gamma + d(x^\beta i_{e^\alpha} e^\gamma) - (\alpha \leftrightarrow \beta) = -\eta^{\beta\gamma}\, dx^\alpha + \eta^{\alpha\gamma}\, dx^\beta.$$

The $A$ of (2.1.10; 2) becomes $A^{\gamma\sigma} = \eta^{\alpha\gamma}\eta^{\beta\sigma} - \eta^{\beta\gamma}\eta^{\alpha\sigma}$, and it satisfies the condition of antisymmetry that characterizes Killing vector fields. Because the interior product is linear, the $\mathscr{T}_v$ of (2.1.8; 4) is simply $x^\beta * \mathscr{T}^\alpha - x^\alpha * \mathscr{T}^\beta$. Remark (2.1.8; 4) means that

$$d(x^\beta * \mathscr{T}^\alpha - x^\alpha * \mathscr{T}^\beta) = x^\beta (i_{e^\alpha} F) \wedge *J - x^\alpha (i_{e^\beta} F) \wedge *J,$$

which implies with (1.3.24) that $dx^\beta \wedge *\mathscr{T}^\alpha - dx^\alpha \wedge *\mathscr{T}^\beta = 0$. Because $dx_\beta \wedge *dx^\gamma = \delta^\gamma{}_\beta * \mathbf{1}$, the energy-momentum tensor must therefore be symmetric; i.e., $T_{\alpha\beta} = T_{\beta\alpha}$. Although (1.3.22) is symmetric, the canonical energy-momentum tensor (2.1.7) is not; to be sure, the canonical energy-momentum tensor is also conserved if $v = x^\beta e^\alpha - x^\alpha e^\beta$, but $x^\alpha$ can not simply be factored out of the gauge-dependent term $di_v A$, which contains $dv$, and the conserved quantity is not $x^\alpha \mathscr{T}^\beta - x^\beta \mathscr{T}^\alpha$.

3. $v = x_\alpha e^\alpha$ generates a dilatation, $L_v e^\gamma = e^\gamma$; and if $F \in E_2(\mathbb{R}^4)$, then according to (2.1.10; 4) $L_v * F = *L_v F$. Then $\mathscr{T}^\alpha$ can be used to formulate a new conservation theorem.

$$d(x_\alpha * \mathscr{T}^\alpha) = x_\alpha i_{e^\alpha} F \wedge *J.$$

From this and (1.3.24) we conclude that $dx \wedge *\mathscr{T}^\alpha = T_\alpha{}^\alpha * \mathbf{1} = 0$. We again observe that (1.3.22) satisfies this equation, whereas the canonical energy-momentum tensor (2.1.7) does not.

4. The conformal transformation $L_v g = 2x^\beta g$ is generated by $v = x^\beta x_\alpha e^\alpha - \frac{1}{2}x^\alpha x_\alpha e^\beta$:

$$L_v e^\gamma = x^\beta e^\gamma + x^\gamma e^\beta - \eta^{\gamma\beta} x_\alpha e^\alpha,$$

and the last two terms cancel out in the expression for $L_v g$. According to Remark (2.1.10; 4), it is again true that $L_v {}^*F = {}^*L_v F$ for all $F \in E_2(\mathbb{R}^4)$, and $d{}^*\mathcal{T}_v$ is

$$d(x^\beta x_\alpha {}^*\mathcal{T}^\alpha - \tfrac{1}{2}x^2 {}^*T^\beta) = ((x^\beta x_\alpha i_{e^\alpha} - \tfrac{1}{2}x^2 i_{e^\beta})F) \wedge {}^*J.$$

The resultant equation

$$0 = x_\alpha \, dx^\beta \wedge {}^*\mathcal{T}^\alpha - x_\alpha \, dx^\alpha \wedge {}^*\mathcal{T}^\beta + x^\beta \, dx_\alpha \wedge {}^*\mathcal{T}^\alpha$$

contains no new information, because the final term vanishes as in Example 3 and the first two vanish as in Example 2.

**Remarks** (2.1.12)

1. The vector fields generating conformal transformations (2.1.11: 4) are not complete (Problem 6); these transformations are not diffeomorphisms of $\mathbb{R}^4$, as they have singularities. This is not serious for the uses we make of them, as we need only the infinitesimal transformations. If a group of diffeomorphisms is desired, then $\mathbb{R}^4$ must be compactified by the addition of points at infinity [17].
2. The Lagrangian $d\Phi \wedge {}^*d\Phi$ for a massless scalar field $\Phi \in E_0$ has no intrinsic length, but even so it fails to be invariant under the dilatation (2.1.11; 3). As a consequence ${T^\alpha}_\alpha \neq 0$ for the energy-momentum tensor of a scalar field even if $m = 0$ (Problem 4).

**The Properties of the Energy-Momentum Tensor** (2.1.13)

Let $e^\alpha$ be an orthogonal basis of a pseudo-Riemannian space ($g = -e^0 \otimes e^0 + \sum_{j=1}^3 e^j \otimes e^j$), and let ${}^*\mathcal{T}^\alpha = \frac{1}{2}(i_{e^\alpha} F \wedge {}^*F - F \wedge i_{e^\alpha} {}^*F) = {T^\alpha}_\beta {}^*e^\beta$. Because of the component representation

$$T_{00} = \tfrac{1}{2}(|\mathbf{E}|^2 + |\mathbf{B}|^2), \qquad {T^j}_0 \equiv \mathcal{S}^j, \qquad j = 1, 2, 3;$$

where

$$\mathcal{S} = [\mathbf{E} \times \mathbf{B}] \text{ is } \textbf{Poynting's vector},$$

we find that for all $x$,

(a) $T_{00}(x) \geq 0$, and $= 0$ only if $F(x) = 0$; and
(b) $\|\mathcal{S}(x)\| \leq T_{00}(x)$.

**Remarks** (2.1.14)

1. If (a) holds in every Lorentz system, then (b) follows. A Lorentz transformation treats $P^\alpha = \int {}^*\mathscr{T}^\alpha$ as a vector (1.3.38; 5), and the equation $P^0 \geq 0$ would be violated by the transformation $P^0 \to (P^0 - \mathbf{v} \cdot \mathbf{P})/\sqrt{1 - v^2}$ unless $\|P\| \leq P^0$.
2. In the orthogonal basis, $\mathscr{S}^j = T^j{}_0 = T_0{}^j$ doubles as a momentum density and as the rate of energy flow: The change in the total energy can be written as

$$\delta\mathscr{T}_0 = \frac{\partial}{\partial t} T_0{}^0 + \frac{\partial}{\partial x^j} T_0{}^j.$$

   The physical interpretation of (b) is that electromagnetic energy can never be transmitted faster than light.
3. The positivity of the energy is expressed mathematically as follows: Let $N_3$ be a spacelike submanifold, and define * with the restriction of $g$ to $N_3$. Then * converts a 3-form into a numerical function, and the positivity of the energy means that ${}^*(T^0{}_{|N_3}) \geq 0$.
4. The signs in (2.1.13) arise from the signature of the metric, and thus depend on the relationship to the standard basis. However, the equations $T^{\alpha\beta} = T^{\beta\alpha}$ and $T^\alpha{}_\alpha = 0$, which follow from (2.1.11; 2) and (2.1.11; 3), hold in any basis on account of the transformation law (1.3.26).

Since the existence of a limiting speed of energy transport follows solely from the structure of the energy-momentum tensor, it is possible to prove the uniqueness of the solution of the Cauchy problem without further analysis of the field equations. Nonuniqueness would contradict the causal propagation of the field at the speed of light. To make this impression mathematically precise, we make some intuitively reasonable

**Definitions** (2.1.15)

(a) A continuous mapping of an interval $I \to M$ is called a **causal curve** iff no two of its points can be connected by a spacelike curve. It is said to be **nonextensible** iff it is not a proper subset of a larger causal curve.

(b) Let $M$ be orientable in time, that is, the forward and backward time directions can be defined smoothly over all of $M$. The **future** (respectively **past**) **domain of influence** $D^+(N)$ (resp. $D^-(N)$) of a spacelike hypersurface $N$ is the set of all points $p$ of $M$ for which all nonextensible, causal curves through $p$ oriented toward the past (resp. future) intersect

$N$. $D(N) \equiv D^+(N) \cup D^-(N)$ is called the **domain of influence** of $N$, and if $D(N) = M$ then $N$ is a **Cauchy surface**

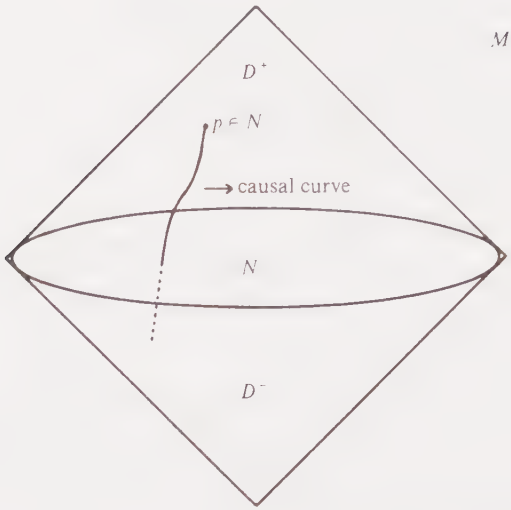



Figure 16 The domain of influence of a hypersurface $N \subset M$

The uniqueness of the Cauchy problem is the statement that $F$ is uniquely determined on $D^{\pm}(N_3)$ by $F_{|N_3}$, ${}^*F_{|N_3}$, and $J$. If $F_{|D^{\pm}(N_3)} \neq 0$ but $J = F_{|N_3} = {}^*F_{|N_3} = 0$, then $F$ would have to propagate from somewhere outside $D^{\mp}(N_3)$ to within $D^{\pm}(N_3)$ without encountering $N_3$, which would be possible only if the propagation speed were greater than the speed of light. Since there exist complicated manifolds $M$ for which the shape of $D^{\pm}$ is confusingly tortuous, let us be contented with the

**Uniqueness of the Cauchy Problem in Minkowski Space** (2.1.16)

*Let $N$ be a three-dimensional, compact, spacelike submanifold of $(\mathbb{R}^4, \eta)$, and suppose that $F_1$ and $F_2$ are two continuous solutions of Maxwell's equations, $dF_1 = dF_2 = 0$, and $\delta F_1 = \delta F_2 = J$. If $F_{1|N} = F_{2|N}$ and ${}^*F_{1|N} = {}^*F_{2|N}$, then $F_1$ and $F_2$ are also equal throughout the interiors of $D^{\pm}(N)$: $F_{1|\operatorname{Int} D^{\pm}(N)} = F_{2|\operatorname{Int} D^{\pm}(N)}$.*

**Proof**

Let $x \in \operatorname{Int} D^+(N)$. Then there exist $\varepsilon > 0$ and $\bar{x} \in D^+(N)$ such that

$$x \in N' \equiv \{y \in \mathbb{R}^4 : |\bar{\mathbf{x}} - \mathbf{y}|^2 - (\bar{x}^0 - y^0)^2 = -\varepsilon^2,\ y^0 \leq \bar{x}^0\}$$
$$\subset \{y \in \mathbb{R}^4 : |\bar{\mathbf{x}} - \mathbf{y}|^2 \leq (\bar{x}^0 - y^0)^2\}.$$

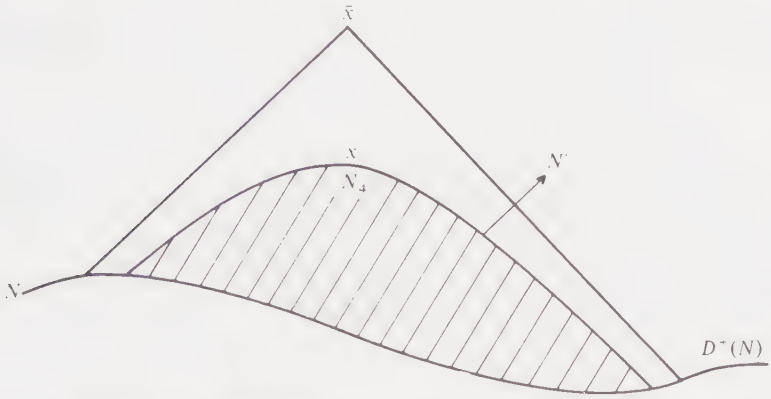



Figure 17 The region of integration used to prove the uniqueness of the Cauchy problem

Hence

$$(N' \cap D^+(N)) \cup \{y \in N : |\bar{\mathbf{x}} - \mathbf{y}|^2 - (\bar{x}^0 - y^0)^2 \leq -\varepsilon^2\} = \partial N_4,$$

where $N_4$ is a compact, four-dimensional submanifold (see Figure 17). If $F = F_1 - F_2$, then $dF = \delta F = 0$ and $F_{|N} = {}^*F_{|N} = 0$. The $\mathscr{T}^\alpha$ formed from $F$ therefore satisfies $\delta \mathscr{T}^\alpha = 0$ and $\mathscr{T}^\alpha{}_{|N} = 0$, and so $0 = \int_{N_4} d^* \mathscr{T}^0 = \int_{N' \cap D^+(N)} {}^* \mathscr{T}^0$. Since $N'$ is spacelike, $\mathscr{T}^0$ is a nonnegative measure on $N'$, and the vanishing of the integral implies that $\mathscr{T}^0{}_{|N' \cap D^+(N)}$ is zero almost everywhere. Since it is a continuous function, $\mathscr{T}^0{}_{|N' \cap D^+(N)} = 0$; and because of (2.1.13) (b), $F_{|N' \cap D^+(N)} = 0$, and in particular $F_{|x} = 0$. The proof for $D^-(N)$ is similar. □

**Remarks** (2.1.17)

1. Formula (1.2.36) proves uniqueness by explicit construction, but it assumes the existence of the distributions $G_{\bar{x}}$, $dG_{x|N}$, and ${}^*dG_{x|N}$. The advantage of the proof given above is that it can be extended to more general manifolds without the necessity of facing the difficult question of the existence of the Green functions.
2. Positivity of the energy is sufficient but not necessary for the uniqueness of the Cauchy problem, which can also be proved (see §3.2), for instance, for the scalar field of Problem 4 with $m^2$ replaced with $-m^2$, although $T^0{}_0$ is then not positive.
3. The electromagnetic current $J$ has no positivity property, and there is no proposition like (2.1.16) for it. Because charges can cancel out, it could happen that $J \equiv 0$ for all $t < t_0$ and $\neq 0$ for all $t > t_0$. More specifically, nothing in Maxwell's equations prevents a charge from moving faster than light.

**Problems** (2.1.18)

1. Show that $(L_v\Phi_j) \wedge (\partial\mathscr{L}/\partial\, d\Phi_j) - i_v\mathscr{L} = di_v G$, if $\mathscr{L} = dG(\Phi)$.

2. Show that if $v$ is a Killing vector, then $L_v{}^* = {}^*L_v$. (Use (1.2.26; 2) and (2.1.10; 2).)

3. The action $W = \int ds[-\dot{z}^\alpha(s)\dot{z}_\alpha(s)]^{1/2}$ is left invariant not only by $s \to s +$ constant, but even by $s \to f(s)$, for $f$ such that $0 < df/ds < \infty$. What is the resultant conserved quantity?

4. Calculate $\mathscr{T}^\alpha$ for $\mathscr{L} = -\frac{1}{2}[d\Phi \wedge {}^*d\Phi + m^2\Phi{}^*\Phi]$, $\Phi \in E_0$, and investigate which of the properties (2.1.11) and (2.1.13) this $\mathscr{T}^\alpha$ has.

5. Calculate the conserved quantity (2.1.5) for $\mathscr{L} = dA \wedge dA$, $A \in E_1$.

6. Integrate the equations $\partial x(t)/\partial t = -a\langle x(t)|x(t)\rangle + 2x(t)\langle a|x(t)\rangle$, which generate the (local) flow of a conformal transformation.

7. Suppose that $\delta(\mathscr{T}^\alpha + t^\alpha) = 0$ and that $\mathscr{T}^\alpha + t^\alpha$ is independent of time in some system and falls off sufficiently in space. Show that in this reference system

$$\int_{0 \le t \le T} dx^j \wedge {}^*(\mathscr{T}^\alpha + t^\alpha) = 0,$$

for $j = 1, 2, 3$, and $\alpha = 0, 1, 2, 3$. (In particular, the "self-stress" terms $\int d^3x\, T_{jj}$ vanish.) Conclude that $\delta(\mathscr{T}^\alpha + t^\alpha)$ can not be zero for the point charge (1.3.25; 2).

**Solutions** (2.1.19)

1. $$\mathscr{L} = d\Phi_j \wedge \frac{\partial G}{\partial \Phi_j}, \qquad (L_v\Phi_j) \wedge \frac{\partial G}{\partial \Phi_j} - i_v\, dG = L_v G - i_v\, dG = di_v G.$$

2. First note that $L_v e^{k_1 \cdots k_p} = \sum_j A^{k_j}{}_k e^{k_1 \cdots k_{j-1} k k_{j+1} \cdots k_p}(-1)^{j+1}$. The identity of (1.2.40; 8) continues to hold when $A^k{}_j$ is substituted for $\omega^k{}_j$, because $A^k{}_j$ has the same antisymmetry. Hence

$$L_v{}^*e^{k_1 \cdots k_p} = \sum_j A^{k_j}{}_k\, {}^*e^{k_1 \cdots k_j k k_{j+1} \cdots k_p}(-1)^{j+1},$$

and consequently

$$\begin{aligned} {}^*L_v(\omega_{k_1 \cdots k_r} e^{k_1 \cdots k_p}) &= (L_v\omega_{k_1 \cdots k_r}){}^*e^{k_1 \cdots k_p} + \cdots \\ &\quad + \omega_{k_1 \cdots k_r}{}^*L_v e^{k_1 \cdots k_p} = L_v(\omega_{k_1 \cdots k_r}{}^*e^{k_1 \cdots k_p}). \end{aligned}$$

3. If $s = \bar{s} + f(\bar{s})$, then as $f \to 0$ with $f = 0$ on the boundary,

$$\begin{aligned} W &= \int d\bar{s}(1 + f')\sqrt{-\dot{z}^\alpha(\bar{s})\dot{z}_\alpha(\bar{s}) - 2f\dot{z}^\alpha(\bar{s})\ddot{z}_\alpha(\bar{s})} \\ &= W + \int d\bar{s}\left(f'\sqrt{-\dot{z}^\alpha(\bar{s})\dot{z}_\alpha(\bar{s})} - f\frac{\dot{z}^\alpha(\bar{s})\ddot{z}_\alpha(\bar{s})}{\sqrt{-\dot{z}^\alpha(\bar{s})\dot{z}^\alpha(\bar{s})}}\right) + O(f^2). \end{aligned}$$

Integrating this by parts leads to the identity

$$\frac{d}{ds}\sqrt{-\dot{z}^\alpha(s)\dot{z}_\alpha(s)} = \frac{-\dot{z}^\alpha(s)\ddot{z}_\alpha(s)}{\sqrt{-\dot{z}^\alpha(s)\dot{z}_\alpha(s)}},$$

which also holds if $\ddot{z} \neq 0$.

4. $\mathcal{T}^\alpha = (i_\alpha d\Phi) \wedge *d\Phi - \frac{1}{2} i_\alpha [d\Phi \wedge *d\Phi + m^2 \Phi * \Phi]$. Then $T_{\alpha\beta} = T_{\beta\alpha}$ (Lorentz invariance), but not $T^\alpha{}_\alpha = 0$, even if $m = 0$ (no dilatation invariance). $T_{00} \geq |\dot{\Phi} \nabla_j \Phi|$ and $T^0{}_j = \dot{\Phi} \nabla_j \Phi$, iff $m^2 \geq 0$.

5. $(L_v A) \wedge \dfrac{\partial \mathscr{L}}{\partial\, dA} - i_v \mathscr{L} = 2(L_v A) \wedge F - i_v (F \wedge F) = 2(d i_v A) \wedge F = 2d(i_v A \wedge F)$.

6. Letting $x \equiv x(0)$, $x(t) = (x - at\langle x|x\rangle)/(1 - 2t\langle x|a\rangle + \langle x|x\rangle\langle a|a\rangle t^2)$, because

$$\frac{\partial x}{\partial t} = \frac{-a\langle x|x\rangle}{1 - 2t\langle x|a\rangle + \langle x|x\rangle\langle a|a\rangle t^2} + 2\,\frac{(x - at\langle x|x\rangle)(\langle x|a\rangle - \langle x|x\rangle\langle a|a\rangle t)}{(1 - 2t\langle x|a\rangle + \langle x|x\rangle\langle a|a\rangle t^2)^2}$$

$$= -a\langle x(t)|x(t)\rangle + 2x(t)\langle a|x(t)\rangle.$$

For any $t > 0$ there exists $x \in \mathbb{R}^4$, namely the $x$ for which

$$1 - 2t\langle a|x\rangle + t^2\langle a|a\rangle\langle x|x\rangle = 0,$$

which gets sent off to infinity.

7. Integrate $x^j\, d*(\mathcal{T}^\alpha + t^\alpha) = dx^j \wedge *(\mathcal{T}^\alpha + t^\alpha) + d(x^j *(\mathcal{T}^\alpha + t^\alpha))$ over

$$N_4 \equiv \{(t, \mathbf{x}) : 0 \leq t \leq T\}.$$

In the rest-frame of the charge nothing depends on time, and only $t^0 \neq 0$. Since $dx_\alpha \wedge *\mathcal{T}^\alpha = 0$ (2.1.11; 3), we obtain the contradiction

$$0 = \sum_{j=1}^{3} \int_{0 \leq t \leq T} dx^j \wedge *\mathcal{T}^j = \int_{0 \leq t \leq T} dx^0 \wedge *\mathcal{T}^0 > 0.$$

## 2.2 The General Solution

*The characteristics of Maxwell's equations are hypersurfaces with lightlike normals. The Green function is easy to construct in Minkowski space* $(\mathbb{R}^4, \eta)$, *and solves the initial-value problem explicitly.*

If the argument of (1.2.28) is applied to the physically interesting case $m = 4$, $p = 2$, one finds that the Minkowski metric

$$\eta = \begin{vmatrix} -1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{vmatrix}$$

contains the $g_+$ of (1.2.28) as its spatial part and also $g_-$ as a spatiotemporal part. Since in the present case $* \circ * = (-)^{p(m-p)+s} = -1$, the combinations of the fields that diagonalize $*$ are $F \pm i*F$, and when these are written explicitly, both real and complex characteristic directions are found, so there is a combination of both of the cases of (1.2.28) (Problem 1). We have remarked that the characteristics may in general be surfaces of discontinuity for the

solutions. If such a hypersurface is specified by $u(x) = 0$, for $u \in E_0(V)$, $V \subset M$, then there should exist solutions that behave locally like $\Theta(u)$. The exterior differential $d\Theta(u) = \delta(u)du$ is then singular at $u = 0$, and if $J$ is a regular function, then the singular contributions to the left sides of the equations $dF = 0$ and $\delta F = J$ must cancel out. These contributions are proportional to $du$, so we are interested in finding solutions for $J = 0$ that depend only on $u$: $F = c_{ij}(u)e^{ij}$. For such solutions the equations say that the exterior and interior products of $F'$ with $du$ must vanish, because $dF = du\, c_{ij}' \wedge e^{ij} \equiv du \wedge F'$, $d{*}F = du \wedge {*}F' = -{*}(i_{du}F')$. This argument leads directly to a

**Condition for the Characteristics** (2.2.1)

*Let $F$ be discontinuous where $u = 0$, but suppose that $J$ is continuous there. Then the equations $dF = 0$ and $\delta F = J$ imply that at $u = 0$*

$$du \wedge F' = 0 \quad \text{and} \quad i_{du}F' = 0;$$

*which are satisfied only if $\langle du|du\rangle = 0$ or $F' = 0$.*

**Proof**

If $du \wedge F' = 0$, then in a local basis using $du$, $F'$ contains $du$ as a factor, that is, $F' = du \wedge f$, where $f \in E_1$ is independent of $du$. The second equation then requires that $i_{du}\, du \wedge f = \langle du|du\rangle f - \langle du|f\rangle du = 0$. Because $f$ and $du$ are independent, we conclude that $\langle du|du\rangle = \langle du|f\rangle = 0$. □

**Remark** (2.2.2)

1. In other words, either the normal to the surface is lightlike, or else there are no discontinuities in $F$.
2. This is only a local statement. Whether $u$ can be defined as a global coordinate depends on the large-scale structure of space-time.
3. A by-product of (2.2.1) is the statement that fields with discontinuities must have a special structure; they are the exterior products of two 1-forms, of which one ($du$) is a null field, and the other ($f$) is orthogonal to it (also in the sense of the metric $\eta$). Both invariants vanish for such fields:

$$*(F \wedge F) = *(du \wedge f \wedge du \wedge f) = 0,$$

and ${*}(F \wedge {*}F) = i_{du}i_f F = (\langle du|f\rangle)^2 - \langle du|du\rangle\langle f|f\rangle = 0$. If the space and time parts are separated, then $du = dt + \mathbf{n}\cdot d\mathbf{x}$ and $f = dt + \mathbf{f}\cdot d\mathbf{x}$, which requires that $|\mathbf{n}|^2 = 1 = (\mathbf{n}\cdot\mathbf{f})$. In terms of the field strengths, $\mathbf{E} = \mathbf{n} - \mathbf{f}$ and $\mathbf{B} = [\mathbf{n} \times \mathbf{f}]$, so this means that $(\mathbf{n}\cdot\mathbf{E}) = (\mathbf{n}\cdot\mathbf{B}) = (\mathbf{E}\cdot\mathbf{B}) = 0 = |\mathbf{E}|^2 - |\mathbf{B}|^2$. The field (1.1.4) was of this form with $u = x - t$.

4. We also note that if $\langle du, du \rangle = 0$, then it is not sufficient to specify $F$ and $^*F$ at $u = 0$ in order to solve the Cauchy problem. It is possible to choose $F$ not identically zero, such that $f(0) = 0$.

5. Moving from one hyperplane $u = \text{const.}$ to the next, the Lie derivative $L_{du}$ in the direction $g\,du$ tangential to $u = \text{const.}$ is determined by Maxwell's equations:

$$L_{du}F = d\iota_{du}F + \iota_{du}dF = 0,$$

$$L_{du}{}^*F = d\iota_{du}{}^*F + \iota_{du}d{}^*F = -\iota_{du}{}^*J.$$

The higher Lie derivatives can similarly be calculated, which determines $F$ everywhere, if $F$ and $^*F$ are specified on some surface that can be translated with $e^{L_{du}}$ so as to cover all of $M$.

After studying these local questions, we will evaluate formula (1.2.36) for Maxwell's equations. This requires the explicit form of the Green function, which can be written down only for the simplest manifolds.

**The Construction of $G_+$ in Minkowski space** $(\mathbb{R}^4, \eta)$ (2.2.3)

If $e^a$ is the natural and orthogonal basis, then according to Equation (1.2.34), $G_+$ has the form

$$G_+ = \tfrac{1}{2}e_{a_1}\cdots e_{a_p}{}^*e^{a_1}\cdots{}^*e^{a_p}D_+(x),$$

for $p = 2$, and where $D_+ \in E_+(\mathbb{R}^4)$ satisfies the equation

$$-D_+{}_{,a}{}^{,a} = \delta^4(x - \bar{x}).$$

The translation-invariance of Minkowski space allows the partial differential equation to be reduced to an ordinary differential equation, by expanding it in a series in the eigenfunctions of the translation operator, i.e., a Fourier series. For the delta-function this expansion is

$$\delta^4(x - \bar{x}) = (2\pi)^{-4}\int_{-\infty}^{+\infty} d^4k\, e^{i\langle k,\, x - \bar{x}\rangle}, \tag{2.2.4}$$

where $x$ and $\bar{x}$ are regarded as the components of vectors, and $\langle\ ,\ \rangle$ stands for the Lorentz scalar product (1.2.14) with $g = \eta$. In Fourier-transformed space the Laplacian $(\cdot)_{,a}{}^{,a}$ produces a factor $ik_a ik^a = -\langle k, k\rangle = -k^2$, so that, finally, $D_+$ has the Fourier integral representation

$$D_+(x) = (2\pi)^{-4}\int_{-\infty}^{+\infty} d^4k\, e^{i\langle k,\, x - \bar{x}\rangle}\, k^{-2}. \tag{2.2.5}$$

**Remarks** (2.2.6)

1. The $k$ integrals do not converge in the classical sense, but as distributions—this is the content of Fourier's theorem. We have unscrupulously interchanged integrals by $k$ with derivatives by $x$. Fortunately, distributions are so agreeable that they put up with manipulations that classical analysis considers criminal.
2. Because of the translation-invariance of Minkowski space, $D_x$ depends only on $x - \bar{x}$.
3. Since $k^2 = (|\mathbf{k}| - k^0)(|\mathbf{k}| + k^0)$, the integrand of (2.2.5) has poles, and it must be decided what to do about them when they are in the integration path. We are not restricted to integrals along the real axis; the analyticity of the integrand of (2.2.4) allows the integration path to be distorted into the complex plane. We use the path denoted as

$$\int_{\frown\frown} d^4k,$$

which passes above the poles in the complex $k$-plane. Other choices of the integration path would produce integrals differing by the contributions of the residues. This nonuniqueness should not be surprising, because the equation in (2.2.3) determines $D_{\bar{x}}$ only up to a solution of the homogeneous equation, and

$$\eta^{\alpha\beta} \frac{\partial^2}{\partial x^\alpha \partial x^\beta} e^{i(\mathbf{k}\cdot\mathbf{x} - k^0 t)} = 0,$$

if $k^0 = \pm|\mathbf{k}|$.

The path shown in Figure 18 is chosen on physical grounds, since the Green function it produces corresponds to physically realistic initial conditions. It is denoted $D^{\text{ret}}(\bar{x} - x)$, and an elementary integration (Problem 2) produces the

**Retarded Green Function** (2.2.7)

$$D^{\text{ret}}(x) = \frac{\delta(r-1)}{4\pi r} = \frac{\delta(x^2)}{2\pi}\,\Theta(t).$$

$$G^{\text{ret}}_{\bar{x}} = \frac{1}{4\pi}\,\bar{e}_{\alpha\beta} \otimes {}^*e^{\alpha\beta}\delta((\bar{x} - x)^2)\Theta(\bar{t} - t),$$

$$r \equiv |\mathbf{x}|, \qquad x^2 \equiv \langle x|x\rangle.$$

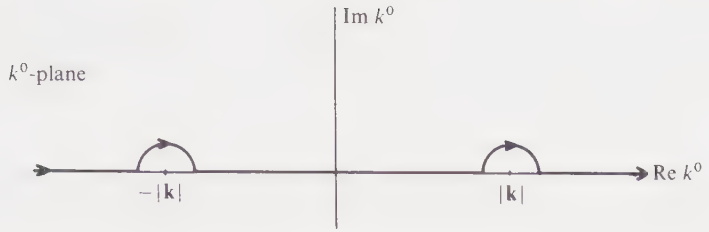



Figure 18 The path of integration for $D^{\text{ret}}$

**Remarks** (2.2.8)

1. For the second form of $D^{\text{ret}}$ we made use of the formula

$$\delta(f(x)) = \sum_i \frac{1}{|f'(x_i)|}\delta(x - x_i), \qquad f(x_i) = 0,$$

to write

$$\delta(x^2)\Theta(t) = \delta((r + t)(r - t))\Theta(t) = \frac{\Theta(t)}{2r}(\delta(r + t) + \delta(r - t))$$

$$= \frac{\delta(r - t)}{2r}\Theta(t).$$

2. The integral (2.2.5) singles out no particular Lorentz system, and thus $D^{\text{ret}}$ depends only on $(x - \bar{x})^2$ and $\Theta(t - \bar{t})$. The preference for one time-direction enters through the choice of the integration path; the path $\neg\cup\cup\neg$ would result in $D^{\text{adv}}(\bar{x} - x) \equiv D^{\text{ret}}(x - \bar{x})$.
3. Because $G_x^{\text{ret}}$ is supported wholly on the negative light-cone of $\bar{x}$ (see Figure 19), the integration over infinite space-time regions is justified painlessly.

The means by which the explicit forms of the field strengths can be calculated from (1.2.36) are now ready. The field is the sum of two integrals, one

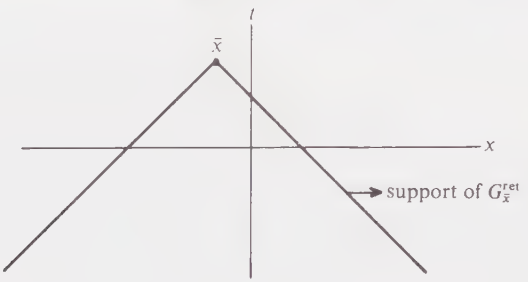



Figure 19 The support property of the retarded Green function

over $N$ and one over $\partial N$, which we call $F^{\text{ret}}$ and $F^{\text{boundary}}$, $F^{\text{ret}}$ depends only on $J$, and $F^{\text{boundary}}$ depends only on the boundary values of the field.

We first consider $F^{\text{ret}}$, which can be calculated with the aid of the formula

$$e^\gamma \wedge {}^*e^{\alpha\beta} \wedge e^\sigma = e^\gamma \wedge e^\sigma \wedge {}^*e^{\alpha\beta} = (\eta^{\sigma\beta}\eta^{\gamma\alpha} - \eta^{\sigma\alpha}\eta^{\gamma\beta}){}^*\mathbf{1}$$

(cf. (1.2.18) (b)) and the elementary rules of calculation (${}^*\mathbf{1} \to d^4x$, $(\partial/\partial x)D^{\text{ret}} = -(\partial/\partial\bar{x})D^{\text{ret}}$):

$$\begin{aligned} F^{\text{ret}}(\bar{x}) &= \int_N dG^{\text{ret}}_{\bar{x}} \wedge J = \tfrac{1}{2}\bar{e}_{\alpha\beta} \int_N \frac{\partial}{\partial x^\gamma} D^{\text{ret}}(\bar{x} - x) e^\gamma \wedge {}^*e^{\alpha\beta} \wedge e^\sigma J_\sigma(x) \\ &= \bar{e}^{\alpha\beta} \int_N d^4x\, J_\alpha(x) \frac{\partial}{\partial \bar{x}^\beta} D^{\text{ret}}(\bar{x} - x) = dA^{\text{ret}}(\bar{x}), \qquad (2.2.9) \\ A^{\text{ret}}(\bar{x}) &\equiv -\bar{e}^\alpha \int_N d^4x\, D^{\text{ret}}(\bar{x} - x) J_\alpha(x). \end{aligned}$$

**Remarks** (2.2.10)

1. $F^{\text{ret}}$ is precisely the exterior differential of a vector potential $A^{\text{ret}}$. It is common to solve for $F$ by first setting $F = dA$, and then using $\delta F = J$ to determine $A$. Equation (1.2.36) with $p = 1$ and $A$ in place of $F$ shows that $A_{|\bar{x}}$ does not depend only on $A_{|\partial N}$, ${}^*dA_{|\partial N}$, and $J_{|N}$, but also on $\delta A_{|N}$. If we rewrite $\int_N \delta G_{\bar{x}} \wedge dA$ as $\int_N G_{\bar{x}} \wedge \delta\, dA + \int_{\partial N} \cdots = \int_N G_{\bar{x}} \wedge J + \int_{\partial N} \ldots$, then there are three contributions to $A(\bar{x})$, viz., $\int_N G_{\bar{x}} \wedge J$ as in (2.2.9), $\int_N dG_x \wedge \delta A$, and $\int_{\partial N} \ldots$. Gauge invariance makes it impossible to fix $A$ in terms of $J$ and the boundary values of $A$ and $dA$: the equations leave open the possibility of a contribution from $d\Lambda$. The solution of the boundary-value problem is unique if we impose the additional condition that $\delta A = 0$ (the **Lorentz gauge**).
2. The $A^{\text{ret}}$ of (2.2.9) satisfies

$$\delta A^{\text{ret}}(\bar{x}) = -\int_{\partial N} D^{\text{ret}}(\bar{x} - x){}^*J(x) \equiv j(\bar{x}), \qquad j \in E_0(N),$$

(Problem 4), so $F^{\text{ret}}$ satisfies the equations

$$dF^{\text{ret}} = 0,$$

$$\delta F^{\text{ret}} = \delta\, d\mathrm{A}^{\text{ret}} = \Delta A^{\text{ret}} - d\delta A^{\text{ret}} = -\int \Delta G^{\text{ret}} \wedge J - dj = J - dj.$$

3. Since the support of $G_x$ is on the negative light-cone of $\bar{x}$, we are not required to choose $N$ compact. If $N$ is taken as $\{(t, x): t_0 \le t \le t_1\}$, then, by (2.2.7), the integral for $A^{\text{ret}}$ reads more explicitly

$$A^{\text{ret}}_\alpha(\bar{x}) = -\int_{|\mathbf{x} - \bar{\mathbf{x}}| < \bar{t} - t_0} \frac{d^3x}{4\pi|\mathbf{x} - \bar{\mathbf{x}}|} J_\alpha(\mathbf{x}, \bar{t} - |\mathbf{x} - \bar{\mathbf{x}}|).$$

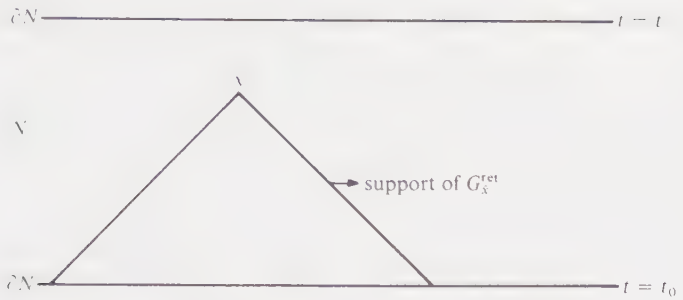



Figure 20 The region of integration for $F^{\text{ret}}(\bar{x})$

This integral always converges for bounded $J_x$, even if $J$ does not have spatially compact support (Figure 20).

4. Although the use of a different Green function, say $D^{\text{adv}}$, does not change $F(\bar{x})$, it does change the integrals $\int_N$ and $\int_{\partial N}$. For example, with the $N$ of Remark 3, $F^{\text{boundary}}$ has contributions only from $t = t_0$, while only the boundary values at $t = t_1$ would show up in $D^{\text{adv}}$. If we want to know $F$ in the half-space later than $t = t_0$, then $D^{\text{ret}}$ is more useful than $D^{\text{adv}}$, because we can let the upper (later) part of $N$ go to infinity. The limit $t_1 \to \infty$ would not necessarily exist for $D^{\text{adv}}$: If we insisted that $F_{|t_0} = 0$, then without the upper boundary term all that would remain of $F$ would be $\int_{t>t_0} dG^{\text{adv}} \wedge J$, which does not vanish at $t = t_0$, and thus requires that there be a contribution to $\int_{\partial N}$ at $t = t_1$. Since the equations are invariant under time-reflection $t \to -t$, the appropriate Green function for the time-reversed question would be $D^{\text{adv}}$.
5. With the aid of Green functions with $p = 3$, we can write (2.2.9) in coordinate-free notation as

$$A^{\text{ret}} = \int_N G_{\bar{x}}^{\text{ret}} \wedge J$$

In order to study $F^{\text{boundary}}$, we use (2.2.7) to write it more explicitly as

$$\begin{aligned} F^{\text{boundary}}(\bar{x}) &= -\int_{\partial N} (\delta G_{\bar{x}} \wedge F - {}^*dG_{\bar{x}} \wedge {}^*F) \\ &= \bar{e}^{\alpha\beta} \int_{\partial N} \Big[ \frac{\partial}{\partial \bar{x}^\alpha} D^{\text{ret}}(\bar{x} - x) F_{\gamma\beta}(x) {}^*e^\gamma - \frac{\partial}{\partial \bar{x}^\gamma} D^{\text{ret}}(\bar{x} - x) F_\alpha{}^\gamma(x) {}^*e_\beta \\ &\quad + \frac{1}{2} \frac{\partial}{\partial \bar{x}^\alpha} D^{\text{ret}}(\bar{x} - x) F_{\alpha\beta}(x) {}^*e^\gamma \Big] \end{aligned} \tag{2.2.11}$$

(Problem 5).

**Remarks** (2.2.12)

1. As in (1.2.38; 1), we see that the boundary $\partial N$ is not arbitrary. If it contained part of the light-cone of a point $\bar{x} \in N$, i.e., the set specified by the equation $t(x) = \bar{t} - |x - \bar{x}|$, then the integral over $\partial N$ would diverge, as

$$\int_{\partial N} D^{\text{ret}}(\bar{x} - x)^* e^0 \cong \int \frac{d^3x}{|x - \bar{x}|}\, \delta(t(\mathbf{x}) - \bar{t} - |\mathbf{x} - \bar{\mathbf{x}}|) \sim \int \delta(0).$$

2. If $N$ is bounded by two Cauchy surfaces (1.2.29) (c), then on the earlier surface $F^{\text{boundary}}$ takes on the boundary values of $F$, and $F^{\text{ret}}$ goes to zero. The first of these claims is proved as follows: If $N$ is given in Cartesian coordinates as $t \geq t(x)$, then†

$$^*e^{\alpha}{}_{|\partial N} = dx^1 \wedge dx^2 \wedge dx^3 \left(1, -\frac{\partial t}{\partial x^1}, \frac{\partial t}{\partial x^2}, -\frac{\partial t}{\partial x^3}\right).$$

The Lorentz system may be chosen so that $\bar{x}$ is the origin and $\partial t(0)/\partial x^j = 0$. Because

$$\lim_{\bar{t} \downarrow 0} \int d^3x\, f(x) \frac{\partial}{\partial \bar{x}^{\beta}} \frac{\delta(\bar{t} - t - r)}{4\pi r}\bigg|_{t=0} = \delta^0{}_{\beta} f(0)$$

(see Problem 6),

$$\lim_{\bar{x} \downarrow 0} \int_{\partial N} {}^*e^{\alpha} \frac{\partial}{\partial \bar{x}^{\beta}} D^{\text{ret}}(\bar{x} - x) f(x) = \delta^{\alpha}{}_0 \delta^0{}_{\beta} f(0).$$

If this is substituted into (2.2.11), the two first terms cancel out as $\bar{x} \downarrow 0$, leaving only $F(0)$. On the other hand

$$\lim_{\bar{x} \downarrow 0} F^{\text{ret}}(\bar{x}) = 0,$$

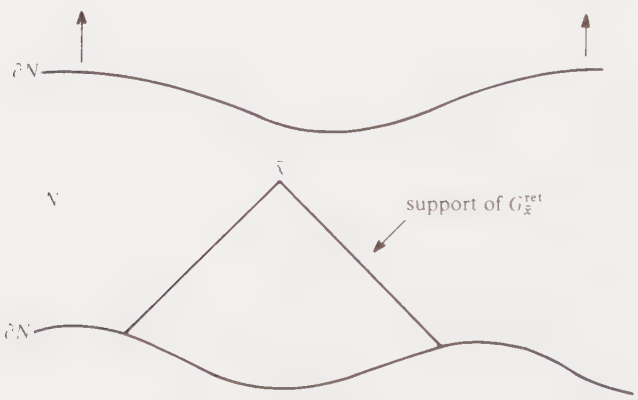



Figure 21 The displacement of the later boundary of $N$

† $^*e^0 = -e^{123}$, but $\partial N$, as the lower side of $N$, has a negative orientation.

because the integration region shrinks to zero (see (2.2.11; 3)), so that if $J$ is bounded, nothing is left over.

3. In Problem 7 it is shown, using the explicit form (2.2.11), that in fact $dF^{\text{boundary}} = 0$ and $\delta F^{\text{boundary}} = dj$, because $d(F_{|\partial N}) = 0$ and $d(*F_{|\partial N}) = -*J$ (cf. (2.2.10; 2)).
4. Although the boundary values normally propagate by $D^{\text{ret}}$ along whole light-cones, particular $F_{|\partial N}$ and $*F_{|\partial N}$ can be found for which the propagation is approximately along a light ray. In this somewhat vague approximation, the problem of how light propagates reduces to the easier problem of the motion of a massless particle. We made use of a similar simplification in (I: 5.7.15) to calculate the bending of light by a gravitational field.
5. By using $D^{\text{ret}}$, one can take $N$ as the whole half-space later than some Cauchy surface, without changing $F^{\text{ret}}$ or $F^{\text{boundary}}$, if the upper boundary is displaced to $t \to +\infty$ (Figure 21).

In our verification that $d(F^{\text{ret}} + F^{\text{boundary}}) = 0$ and $d(*F^{\text{ret}} + *F^{\text{boundary}}) = -*J$, i.e.,

$$\lim_{\bar{x} \downarrow x} F^{\text{ret}}(\bar{x}) = \lim_{\bar{x} \downarrow x} *F^{\text{ret}}(\bar{x}) = 0, \qquad \lim_{\bar{x} \downarrow x} F^{\text{boundary}}(\bar{x}) = F(x),$$

$$\lim_{\bar{x} \downarrow x} *F^{\text{boundary}}(\bar{x}) = *F(x) \quad \forall x \in \partial N,$$

we needed to know only that $d(F_{|\partial N}) = 0$ and $d(*F_{|\partial N}) = -*J_{|\partial N}$. Whereas to use (1.2.36) it is a priori necessary to assume the existence of a solution $F$, the explicit construction demonstrates the

**Existence of the Solution to $dF = 0$, $\delta F = J$, for Given Initial Values** (2.2.13)

Let $N$ be the half-space later than a Cauchy surface $\partial N$ in $(\mathbb{R}^4, \eta)$, and suppose that $H \in E_2(\partial N)$, $K \in E_2(\partial N)$, and $J \in E_1(N)$ satisfy the equations

$$dH = 0, \qquad dK = -*J_{|\partial N}.$$

Then

$$F(\bar{x}) = \int_N dG_{\bar{x}} \wedge J - \int_{\partial N} [\delta G_{\bar{x}} \wedge H - *dG_{\bar{x}} \wedge K]$$

is a solution of $dF = 0$ and $\delta F = J$ on $N$, for which $F$ approaches $H$ and $*F$ approaches $K$ as $x$ approaches $\partial N$.

**Remarks** (2.2.14)

1. On a Cauchy surface $\partial N$, $F_{|\partial N}$ and $*F_{|\partial N}$ are linearly independent, so the boundary values $H$ and $K$ do not depend on each other. If $N$ had characteristic directions, then the above construction would not work, because $\delta G_{\bar{x}}$ and $*dG_{\bar{x}}$ are not distributions on lightlike surfaces.

2. $H$ and $K$ satisfy the restriction of Maxwell's equations to a spacelike hypersurface; $\nabla \cdot \mathbf{H} = 0$ and $\nabla \cdot \mathbf{K} = \rho$. Thus they may be taken as $\mathbf{H} = \nabla \times \mathbf{v}_1$ and $\mathbf{K} = \nabla \times \mathbf{v}_2 - \nabla \int \rho(x)/|\mathbf{x} - \bar{\mathbf{x}}|$, for arbitrary $\mathbf{v}_1$ and $\mathbf{v}_2$.

Often $J$ is sufficiently localized that even $\int_{\mathbb{R}^4} dG_x \wedge J$ converges. Then the remaining boundary in (1.2.36) can be taken to $t_0 = -\infty$. If $F^{\text{ret}}$ is to converge in this case, $F^{\text{boundary}}$ must also converge, and in fact it approaches a solution of the free equation ($J = 0$), because $F^{\text{ret}}$ then solves Maxwell's equations. This incoming field will be denoted by $F^{\text{in}}$, and in the time-reversed situation, using $D^{\text{adv}}$ from (2.2.8; 2), it will be called $F^{\text{out}}$.

**Definition of the Asymptotic Fields** (2.2.15)

$$F = F^{\text{in}} + F^{\text{ret}} = F^{\text{out}} + F^{\text{adv}},$$

if the integrals

$$F^{\text{ret}}(\bar{x}) = \int_{\mathbb{R}^4} dG^{\text{ret}}_{\bar{x}} \wedge J \quad \text{and} \quad F^{\text{adv}}(\bar{x}) = \int_{\mathbb{R}^4} dG^{\text{adv}}_{x} \wedge J$$

exist; in which case

$$dF^{\text{ret}} = dF^{\text{adv}} = dF^{\text{in}} = dF^{\text{out}} = \delta F^{\text{in}} = \delta F^{\text{out}} = 0, \qquad \delta F^{\text{ret}} = \delta F^{\text{adv}} = J.$$

**Remarks** (2.2.16)

1. The existence of $F^{\text{ret}}$ is guaranteed if $J$ is localized well enough in space and does not evolve too nearly along a light-cone. We shall learn from the hyperbolic motion in §2.3 that it is possible for $F^{\text{ret}}$ to converge pointwise while $\delta F^{\text{ret}} \neq J$, because the limit as $t_0 \to -\infty$ does not exist in the appropriate topology.
2. If there exists a $T$ such that $J = 0 \,\forall t < T$, which is, of course, possible only if the total charge $Q = 0$, then $F = F^{\text{in}} \,\forall\, t < T$. Roughly speaking, $F^{\text{in}}$ is the field that existed before $J$ was switched on, and $F^{\text{out}}$ is the field that remains after $J$ has been switched off.

If $N = \mathbb{R}^4$, then $F^{\text{ret}}$ is the differential of the

**Liénard–Wiechert Potentials** (2.2.17)

$$F^{\text{ret}} = dA^{\text{ret}},$$

where

$$A^{\text{ret}}_x(\bar{x}) = -\int \frac{d^3x\, J_x(\bar{t} - |\bar{\mathbf{x}} - \mathbf{x}|, x)}{4\pi |\bar{\mathbf{x}} - \mathbf{x}|}.$$

**The Static Case** (2.2.18)

If $J$ is independent of time, then

$$A_\alpha^{\text{ret}}(\bar{\mathbf{x}}) = -\int \frac{d^3x\, J_\alpha(x)}{4\pi|\bar{\mathbf{x}} - \mathbf{x}|},$$

and, in particular, the Coulomb potential

$$V(\bar{\mathbf{x}}) = A_0^{\text{ret}}(\bar{\mathbf{x}}) = \int \frac{d^3x\, \rho(x)}{4\pi|\bar{\mathbf{x}} - \mathbf{x}|}, \qquad \rho(x) \equiv J^0(x) = -J_0(x).$$

A necessary condition for the existence of this potential is that $\rho$ fall off faster than $1/r^2$.

We close this section with a different application of (2.2.15), to rewrite Equation (1.3.23) for the work expended. To this end we assume that $J$ is sufficiently well localized in space so that $N$ may be set to $\mathbb{R}^4$. Then

$$P^\alpha(T) - P^\alpha(-T) = \int_{-T<t\le T} (i^\alpha F^{\text{in}} + i^\alpha F^{\text{ret}}) \wedge {}^*J. \tag{2.2.19}$$

The energy and momentum are thus affected partly by the incoming field and partly by the field caused by $J$ alone. The latter contribution is explicitly

$$\int_{-T\le t\le T} i^\alpha F^{\text{ret}} \wedge {}^*J = \int_{-T}^{T} d^4\bar{x} \int_{-\infty}^{\infty} d^4x\, J^\beta(\bar{x}) \times (J^\alpha(x) \frac{\partial}{\partial \bar{x}^\beta} D^{\text{ret}}(\bar{x} - x) - J_\beta(x) \frac{\partial}{\partial \bar{x}_\alpha} D^{\text{ret}}(\bar{x} - x)). \tag{2.2.20}$$

If $J$ has compact support ($Q = 0$), then we can take the limit $T \to \infty$. By integration by parts, the first term on the right side of (2.2.20) contributes zero, because $\delta J = 0$. By symmetrization $x \leftrightarrow \bar{x}$, in the second term

$$\frac{\partial}{\partial \bar{x}^\alpha} D^{\text{ret}}(\bar{x} - x)$$

is changed to

$$\frac{1}{2}\frac{\partial}{\partial \bar{x}^\alpha}(D^{\text{ret}}(\bar{x} - x) - D^{\text{adv}}(\bar{x} - x)) = \frac{1}{2}\frac{\partial}{\partial \bar{x}^\alpha} D(\bar{x} - x).$$

Introducing the **radiation field**

$$F^{\text{rad}} = F^{\text{ret}} - F^{\text{adv}} = F^{\text{out}} - F^{\text{in}}$$

allows one to write

$$P_\alpha(\infty) - P_\alpha(-\infty) = \int_{\mathbb{R}^4} (i_\alpha F^{\text{in}} + i_\alpha \tfrac{1}{2} F^{\text{rad}}) \wedge {}^*J,$$

$$-\int i_\alpha F^{\text{rad}} \wedge {}^*J = \int d^4\bar{x}\, d^4x\, J^\beta(\bar{x}) J_\beta(x) \frac{\partial}{\partial \bar{x}^\alpha} D(\bar{x} - x). \tag{2.2.21}$$

The convolution (2.2.21) becomes a product when Fourier-transformed†:

$$\tilde{J}_\beta(k) = \int d^4x\, e^{i\langle k|x\rangle} J_\beta(x) = \tilde{J}_\beta{}^*(-k).$$

Since $D$ has the Fourier integral representation

$$D(x) = i(2\pi)^{-3} \int d^4k\, \delta(k^2)(\Theta(k^0) - \Theta(-k^0)) e^{i\langle k|x\rangle} \tag{2.2.22}$$

(Problem 8), there results an expression for the

**Energy and Momentum Lost by the Radiative Reaction of the Field** (2.2.23)

If $F^{\text{in}} = 0$, $J \in E_1^0(\mathbb{R}^4)$, then

$$P^\alpha(\infty) - P^\alpha(-\infty) = (2\pi)^{-3} \int d^4k\, \Theta(k^0)\delta(k^2)(|\tilde{\mathbf{J}}(k)|^2 - |\tilde{\mathbf{J}}^0(k)|^2) k^\alpha.$$

**Remarks** (2.2.24)

1. Because $\tilde{F}^{\text{ret}}_{\alpha\beta}(k) = (ik_\alpha \tilde{J}_\beta(k) - ik_\beta \tilde{J}_\alpha(k))\, k^{-2}$, the field reflects the frequency distribution of the source. This allows the integrand to be interpreted as the energy and momentum lost to the field with wave vector $k$. In particular, as the time becomes infinite, the factor $\delta(k^2)$ makes the loss go mainly into the free field, characterized by $|\mathbf{k}|^2 = (k^0)^2$.
2. If the sources are strictly periodic, the assumption that $J \in E_1^0$ is violated. This shows up as a delta function in $\tilde{J}$ and a $\delta^2$ in (2.2.23). The physical significance is that a periodic process radiates an infinite amount of energy in an infinite time (cf. (2.3.29; 7)).
3. In Fourier-transformed space, $\delta J = 0$ reads

$$\tilde{\mathbf{J}} \cdot \mathbf{k} = \tilde{J}^0 k^0.$$

If $|\mathbf{k}| = |k^0|$, then $|\mathbf{J}| \geq |J^0|$. Since with $\alpha = 0$ the rest of the integrand is nonnegative, $P^0(\infty) \geq P^0(-\infty)$. This is a consequence of the positivity

† We shall write * on the right side to denote the complex conjugate.

of the energy. If $F^{\text{in}} = 0$, then $P^0(-\infty) = 0$; thus energy can only be released from the current to the field.

**Problems** (2.2.25)

1. Write $F \pm i^*F$ out explicitly, and show that $E_1 = B_2 =$ an arbitrary function of $t + z$, and $E_2 = B_1 = E_3 = B_3 = 0$ solves the equations $dF = \delta F = 0$. Show that the field can be written as $du \wedge f$, where $\langle du | du \rangle = \langle du | f \rangle = 0$.
2. Calculate the integral (2.2.5) for $D^{\text{ret}}$, using the integration path (2.2.6; 3).
3. Use $D^{\text{ret}}$ to find the Green function for Laplace's equation in three dimensions.
4. Calculate $\delta A$ for the $A$ of (2.2.9).
5. Calculate $\delta G_{\bar{x}} \wedge F - {}^*dG_{\bar{x}} \wedge {}^*F$ with the $G_{\bar{x}}$ of (2.2.7) and $F = \frac{1}{2}e^{\alpha\beta}F_{\alpha\beta}$.
6. Show that
$$\lim_{t \to 0} \int d^3x\, f(x) \frac{\partial}{\partial x^\alpha} \frac{\delta(t - r)}{4\pi r} = \delta^0{}_\alpha f(0).$$
7. Write (2.2.11) for $\partial N = (0, \mathbf{x})$ explicitly in terms of the components $\mathbf{E}$ and $\mathbf{B}$ of $F^{\text{boundary}}$, and calculate $E_{k\,k}$ and $B_{k\,k}$.
8. Find the Fourier integral decomposition of the $D$ of (2.2.22).

**Solutions** (2.2.26)

1. $$\begin{aligned} F + i^*F = {} & \tfrac{1}{2}(E_1 + B_2)(dt + dx_3)(dx_1 - i\,dx_2) + \tfrac{1}{2}(E_1 - B_2)(dt - dx_3)(dx_1 + i\,dx_2) \\ & + \tfrac{1}{2}(E_2 + B_3)(dt + dx_1)(dx_2 - i\,dx_3) + \tfrac{1}{2}(E_2 - B_3)(dt - dx_1)(dx_2 + i\,dx_3) \\ & + \tfrac{1}{2}(E_3 + B_1)(dt + dx_2)(dx_3 - i\,dx_1) + \tfrac{1}{2}(E_3 - B_1)(dt - dx_2)(dx_3 + i\,dx_1). \end{aligned}$$
The individual terms of this sum are of the required form.
2. If $t \lesseqgtr 0$ (respectively $\gtreqless 0$), then the path of integration can be closed in the upper (resp. lower) complex $k^0$-plane, so that
$$-\frac{1}{2\pi} \int_{\frown\!\!\!\!\!\!\smile} dk^0 \frac{e^{-ik^0 t}}{(k^0 - k)(k^0 + k)} = \frac{\Theta(t)}{k} \sin kt.$$
There remains
$$(2\pi)^{-3} \int d^3k\, e^{i\mathbf{k}\cdot\mathbf{x}} \frac{\Theta(t)}{k} \sin kt = \frac{\Theta(t)}{2\pi^2 r} \int_0^\infty dk \sin kr \sin kt = \frac{\delta(r - t)}{4\pi r} \Theta(t).$$
3. Because $(\partial^2/\partial t^2) - \Delta) D^{\text{ret}}(x) = \delta^4(x)$, the function $\int_{-\infty}^{\infty} dt\, D^{\text{ret}}(x) = 1/4\pi r$ satisfies the equation $-\Delta(1/4\pi r) = \delta^3(x)$.
4. $$-\delta A^{\text{ret}}(\bar{x}) = \int_N d^4x\, J_\alpha(x) \frac{\partial}{\partial x^\alpha} D^{\text{ret}}(\bar{x} - x) \equiv \int_N d({}^*JD^{\text{ret}}_{\bar{x}}) = \int_{\partial N} {}^*JD^{\text{ret}}_{\bar{x}}.$$

5. Using the abbreviation $D_{,x} = (\partial/\partial x^\alpha)D^{\text{ret}}(\bar{x} - x) = -(\partial/\partial \bar{x}^\alpha)D^{\text{ret}}(\bar{x} - x)$ and Rule (1.2.18)(b),

$$dG_{\bar{x}} = \tfrac{1}{2}\bar{e}_{\alpha\beta}e^\gamma \wedge {}^*e^{\alpha\beta}D_{,\gamma} = -\bar{e}^{\alpha\beta}D_{,\alpha}{}^*e_\beta$$

$$^*dG_{\bar{x}} \wedge {}^*F = -\tfrac{1}{2}\bar{e}^{\alpha\beta}D_{,\alpha}e_\beta \wedge {}^*e^{\sigma\tau}F_{\sigma\tau} = \bar{e}^{\alpha\beta}D_{,\alpha}F_{\beta\sigma}{}^*e^\sigma$$

$$\delta G_{\bar{x}} = \tfrac{1}{2}\bar{e}^{\alpha\beta}{}^*(dD \wedge e_{\alpha\beta}) = \tfrac{1}{2}\bar{e}^{\alpha\beta}D_{,}{}^\gamma{}^*e_{\gamma\alpha\beta}$$

$$\delta G_{\bar{x}} \wedge F = \tfrac{1}{4}\bar{e}_{\alpha\beta}D_{,\gamma}F_{\sigma\tau}e^{\sigma\tau} \wedge {}^*e^{\gamma\alpha\beta} = -\bar{e}_{\alpha\beta}D_{,\gamma}[F^{\alpha\gamma}{}^*e^\beta - \tfrac{1}{2}F^{\alpha\beta}{}^*e^\gamma].$$

If these equations are combined, then

$$-\delta G_{\bar{x}} \wedge F + {}^*dG_{\bar{x}} \wedge {}^*F = \bar{e}^{\alpha\beta}\Big[\frac{\partial}{\partial \bar{x}^\alpha}D(\bar{x} - x)F_{\gamma\beta}(x){}^*e^\gamma - \frac{\partial}{\partial \bar{x}_\gamma}D(\bar{x} - x)F_{\alpha\gamma}(x){}^*e_\beta + \frac{1}{2}\frac{\partial}{\partial \bar{x}_\gamma}D(\bar{x} - x)F_{\alpha\beta}(x){}^*e^\gamma\Big].$$

6. $\alpha = j = 1, 2, 3$: $\displaystyle\int_0^\infty dr\, d\Omega(f(0)r^2 + f_{,k}(0)r^2x_k + \cdots)\frac{-\partial}{\partial x_j}\frac{\delta(t-r)}{4\pi r}$

$$= \int_0^\infty dr\, d\Omega(f(0)2x_j + f_{,k}(0)(2x_jx_k + r^2\delta^j{}_k) + \cdots)\frac{\delta(t-r)}{4\pi r} \to 0 \qquad \text{as } t \to 0.$$

$\alpha = 0$: $\displaystyle\int_0^\infty dr\, d\Omega(f(0)r^2 + f_{,k}(0)r^2x_k + \cdots)\frac{1}{4\pi r}\frac{-\partial}{\partial r}\delta(t-r) = f(0) + O(t) \to f(0)$

as $t \to 0$.

7. $$E_k(\bar{x}) = \int d^3x[\dot{D}^{\text{ret}}(\bar{x} - x)E_k(x) + \varepsilon_{klm}D^{\text{ret}}_{,l}(\bar{x} - x)B_m(x)]$$

$$B_k(\bar{x}) = \int d^3x[\dot{D}^{\text{ret}}(\bar{x} - x)B_k(x) - \varepsilon_{klm}D^{\text{ret}}_{,l}(\bar{x} - x)E_m(x)]$$

$$E_{k,k}(\bar{x}) = \int d^3x\, \dot{D}^{\text{ret}}(\bar{x} - x)E_{k,k}(x) = \frac{\partial}{\partial \bar{t}}\int d^3x\, D^{\text{ret}}(\bar{x} - x)J^0(x)$$

$$B_{k,k}(\bar{x}) = \int d^3x\, \dot{D}^{\text{ret}}(\bar{x} - x)B_{k,k}(x) = 0,$$

since **E** and **B** must satisfy the restriction of Maxwell's equations to $\partial N$.

8. $D(x) = (D^{\text{ret}}(x) - D^{\text{adv}}(x))$

$$= (2\pi)^{-4}\int d^4k\, e^{i\langle k|x\rangle}\lim_{\varepsilon \downarrow 0}\left[\frac{1}{|\mathbf{k}|^2 - (k^0 + i\varepsilon)^2} - \frac{1}{|\mathbf{k}|^2 - (k^0 - i\varepsilon)^2}\right]$$

$$= (2\pi)^{-3}i\int d^4k\, e^{i\langle k|x\rangle}\delta(k^2)[\Theta(k^0) - \Theta(-k^0)].$$

## 2.3 The Field of a Point Charge

*There is an expression in closed form for the field of a point charge undergoing any given motion. It contains all information about the radiation emitted by an accelerating charge.*

Our first application of the formulas derived in the preceding section will be to calculate $A^{\text{ret}}$ and $F^{\text{ret}}$ for the current (1.3.25; 2); $J$ and $D^{\text{ret}}$ will be respectively a four-dimensional and a one-dimensional delta function, which allow the integrals over $d^4x$ and $ds$ to be done. Using the rule $\delta(f(x)) = \sum_i \delta(x - x_i)|f'(x_i)|^{-1}$, where the $x_i$ are the zeroes of $f$, we obtain

$$A_\alpha^{\text{ret}}(\bar{x}) = -\int d^4x\, D^{\text{ret}}(\bar{x} - x)J_\alpha(x) = -e\int_{-\infty}^{\infty} ds\, \dot{z}_\alpha(s)D^{\text{ret}}(\bar{x} - z(s))$$

$$= \frac{-e}{2\pi}\int_{-\infty}^{z^0(s)<\bar{x}^0} ds\, \dot{z}_\alpha(s)\delta((\bar{x} - z(s))^2) = \frac{e}{4\pi}\frac{\dot{z}_\alpha(s_0)}{\langle \dot{z}(s_0)|\bar{x} - z(s_0)\rangle}, \tag{2.3.1}$$

$$(\bar{x} - z(s_0))^2 = 0, \qquad \bar{x}^0 > z^0(s_0).$$

**Remarks** (2.3.2)

1. The negative sign makes (2.3.1) consistent with $\langle \dot{z}(s_0)|\bar{x} - z(s_0)\rangle < 0$.
2. $s_0$ is a function of $\bar{x}$ (see Figure 22).
3. If we consider $\bar{x}_\alpha$, $z_\alpha(s_0(\bar{x}))$, and $\dot{z}_\alpha(s_0(\bar{x}))$ as the components of vector fields $x$, $z$, and $\dot{z}$, then (2.3.1) may be written without indices as

$$A^{\text{ret}} = \frac{e}{4\pi}\frac{\dot{z}}{\langle \dot{z}|x - z\rangle}.$$

After the $x$-integration the coordinate will be called $x$, rather than $\bar{x}$.

The way that $s_0$ depends on $x$ must first be known before $F^{\text{ret}}$ can be calculated from $A^{\text{ret}}$. This dependence can be determined by noting that $x - z \in E_1(\mathbb{R}^4)$ is a null field:

$$0 = \frac{1}{2}\frac{\partial}{\partial x^\beta}(x^\alpha - z^\alpha(s_0(x)))(x_\alpha - z_\alpha(s_0(x))) = x_\beta - z_\beta - \dot{z}^\alpha\frac{\partial s_0}{\partial x^\beta}(x_\alpha - z_\alpha)$$

$$\Rightarrow ds_0 = (x - z)\langle \dot{z}|x - z\rangle^{-1}. \tag{2.3.3}$$

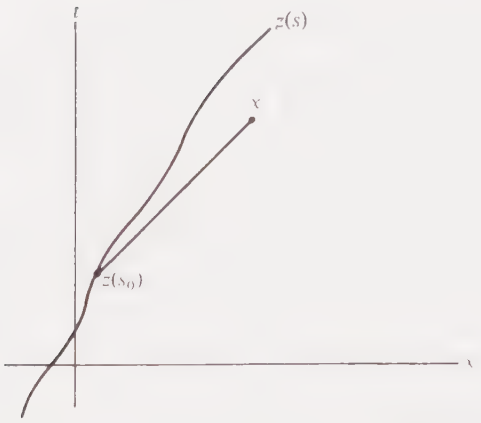



Figure 22 Determination of $z(s_0)$ on the world-line of a charge

The exterior differential of the constituents of $A$ can be calculated from (2.3.3) as

$$d\dot{z} = ds_0 \wedge \ddot{z} = (x - z) \wedge \ddot{z}\langle \dot{z} | x - z\rangle^{-1},$$

$$\frac{\partial}{\partial x^\beta}\dot{z}^\alpha(x_\alpha - z_\alpha) = \ddot{z}^\alpha \frac{\partial s_0}{\partial x^\beta}(x_\alpha - z_\alpha) + \dot{z}_\beta - \dot{z}^\alpha \dot{z}_\alpha \frac{\partial s_0}{\partial x^\beta}$$

$$\Rightarrow d\langle \dot{z} | x - z\rangle = \dot{z} + (x - z)\frac{\langle \ddot{z} | x - z\rangle + 1}{\langle \dot{z} | x - z\rangle}. \tag{2.3.4}$$

where we have used the normalization $\dot{z}^\alpha \dot{z}_\alpha \equiv \dot{z}^2 \equiv \langle \dot{z} | \dot{z}\rangle = -1$ and have also considered $\ddot{z}_\alpha(s_0(x))$ as the components of the vector field $\ddot{z}$. Combining the above formulas produces

**The Retarded Field of a Point Charge** (2.3.5)

$$F^{\text{ret}} = \frac{e}{4\pi}\left(\langle \dot{z} | x - z\rangle^{-1} d\dot{z} + \langle \dot{z} | x - z\rangle^{-2}\dot{z} \wedge d\langle \dot{z} | x - z\rangle\right)$$

$$= \frac{e}{4\pi}\langle \dot{z} | x - z\rangle^{-2}\left(\dot{z}\frac{\langle \ddot{z} | x - z\rangle + 1}{\langle \dot{z} | x - z\rangle} - \ddot{z}\right) \wedge (x - z).$$

**Remarks** (2.3.6)

1. By using the normalization $\dot{z}^2 = -1$ for the world-line $z(s)$ we assume that the charge never reaches the speed of light.

2. The 2-form $F^{\text{ret}}$ is of a special form, as the exterior product of a null vector field $x - z$ with another vector field $\dot{z}\langle \dot{z} | x - z\rangle^{-1}(\langle \ddot{z} | x - z\rangle + 1) - \ddot{z}$. In contrast to the field in (2.2.2; 3) with discontinuities, the interior product of these fields is not 0 but 1.
3. $F^{\text{ret}}$ is the sum of two fields, one of which, $F^{(\ddot{z})}$, contains the terms proportional to $\ddot{z}$, and the other, $F^{(\dot{z})}$, contains only $\dot{z}$. The two fields have different asymptotic behavior, as expected on dimensional grounds; $F^{(\dot{z})}$, the field of the near zone, falls off as a Coulomb field $(1/r^2)$, whereas $F^{(\ddot{z})}$, the field of the far zone, falls off only as $1/r$.

We note some of the special

**Properties of $F^{\text{ret}}$ for a Point Charge** (2.3.7)

(a) $F^{\text{ret}} \wedge F^{\text{ret}} = 0 = F^{(\dot{z})} \wedge F^{(\dot{z})} = F^{(\ddot{z})} \wedge F^{(\ddot{z})}$
(b) $F^{\text{ret}} \wedge (x - z) = 0 = F^{(\dot{z})} \wedge (x - z) = F^{(\ddot{z})} \wedge (x - z)$
(c) $i_{x-z}F^{(\ddot{z})} = {}^*((x - z) \wedge {}^*F^{(\ddot{z})}) = 0$
(d) $F^{(\ddot{z})} \wedge {}^*F^{(\ddot{z})} = 0 = F^{(\dot{z})} \wedge {}^*F^{(\ddot{z})} = F^{\text{ret}} \wedge {}^*F^{(\ddot{z})}$.

**Proof**

(a) This holds for any element of $E_2$ that is an exterior product of two vectors.
(b) $F^{\text{ret}}$, $F^{(\dot{z})}$, and $F^{(\ddot{z})}$ contain $x - z$ as a factor.
(c) The interior product of the two factors of $F^{(\ddot{z})}$ vanishes, and $x - z$ is a null field.
(d) This follows from (c) because of the factor $x - z$. □

**Remarks** (2.3.8)

1. Since $F \wedge F \sim \mathbf{E} \cdot \mathbf{B}$, Property (a) implies that the electric and magnetic fields are always perpendicular to each other, and this is also a property of $F^{(\dot{z})}$ and $F^{(\ddot{z})}$ considered separately. Therefore it requires more than one charge to produce a magnetic field parallel to an electric field.
2. Because ${}^*((x - z) \wedge F) = i_{x-z}{}^*F$, Property (b) implies that

$$(x - z)^\beta ({}^*F^{\text{ret}})_{\alpha\beta} = 0.$$

For $\alpha = 0$, this means that $\mathbf{B}$ is also perpendicular to the 3-vector connecting the reference point (where the field is measured) to the position of the charge at the retarded time $s_0$.
3. If $F^{(\dot{z})}$ is considered separately, then Remark 2 still applies, and because of Property (c) it applies even if $F$ and ${}^*F$, that is, $\mathbf{B}$ and $\mathbf{E}$, are interchanged. Therefore the electric field stemming from $F^{(\dot{z})}$, which dominates in the far zone, is also perpendicular to the 3-vector from the reference point to the position of the charge at the retarded time $s_0$.

4. According to Property (d), because $F \wedge {}^*F \sim |\mathbf{E}|^2 - |\mathbf{B}|^2$, in the far zone $|\mathbf{E}^{(\ddot{z})}| = |\mathbf{B}^{(\ddot{z})}|$, and if $z(s_0) = 0$, then

$$\mathbf{E}^{(\ddot{z})} = \left[\mathbf{B}^{(\ddot{z})} \times \frac{\mathbf{x}}{r}\right] \quad \text{and} \quad \mathbf{B}^{(\ddot{z})} = \left[\frac{\mathbf{x}}{r} \times \mathbf{E}^{(\ddot{z})}\right].$$

The Lagrangian $F^{\text{ret}} \wedge {}^*F^{\text{ret}} = F^{(\dot{z})} \wedge {}^*F^{(\dot{z})}$ has contributions only from the near zone.

**The Field in the Rest-Frame of the Particle** (2.3.9)

Let us choose the Lorentz system in which $z(s_0) = 0$, $\dot{z}(s_0) = (1, 0, 0, 0)$, and $\ddot{z}(s_0) = (0, \ddot{z})$. Then for $x$ in the positive light-cone of $z(s_0)$, the field (2.3.5), written in components, is expressed as

$$\mathbf{E} = \frac{e}{4\pi r}\left\{-\ddot{\mathbf{z}} + \frac{\mathbf{x}}{r^2}(1 + (\mathbf{x} \cdot \ddot{\mathbf{z}}))\right\}$$

$$\mathbf{B} = \frac{e}{4\pi r}\left[\ddot{\mathbf{z}} \times \frac{\mathbf{x}}{r}\right].$$

The magnetic lines of force circle the charge, and the electric lines of force leave the charge initially in the radial direction, and later bend to become perpendicular to **B** and to **x** in the far zone (Figure 23)

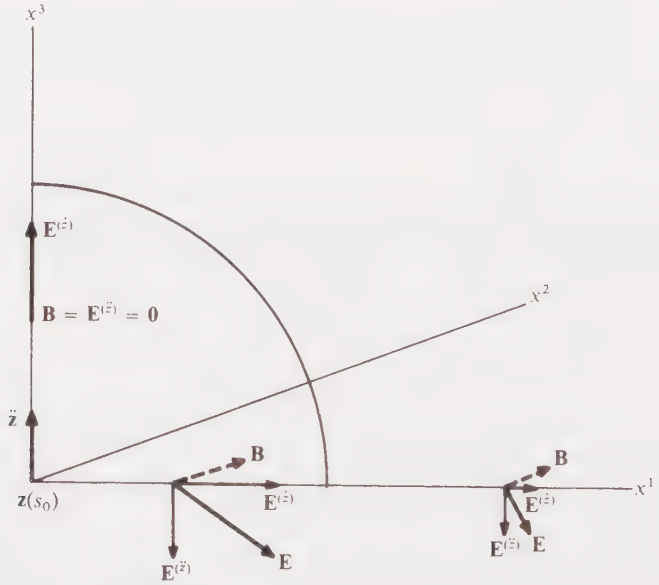



Figure 23 The fields in the near and far zones

**Warning** (2.3.10)

The description in (2.3.9) is not an instantaneous picture of the field, but specifies the field at a given position $\mathbf{x}$ at the time $t = |\mathbf{x}|$. The field in the spacelike section at $t =$ constant does not depend on $\dot{z}$ and $\ddot{z}$ at a single value of $s$, but on a whole segment of the particle's trajectory. It can be written down explicitly only for special kinds of trajectories.

**Examples** (2.3.11)

1. Uniform motion. Let

$$z(s) = \frac{s}{\sqrt{1 - v^2}}(1, v, 0, 0) = \dot{z}s.$$

Then

$$(x - z(s_0))^2 = x^2 - s_0^2 - 2\langle \dot{z} | x \rangle s_0,$$

and thus

$$s_0 = -\langle \dot{z} | z(s_0) \rangle = -\langle \dot{z} | x \rangle - \sqrt{\langle \dot{z} | x \rangle^2 + x^2}.$$

This means that

$$A^{\text{ret}} = \frac{-e}{4\pi} \frac{\dot{z}}{\sqrt{\langle \dot{z} | x \rangle^2 + x^2}}$$

$$= \frac{-e}{4\pi} \frac{(-1, v, 0, 0)}{\sqrt{1 - v^2}} \left[ \left( \frac{x_1 - vt}{\sqrt{1 - v^2}} \right)^2 + x_2^2 + x_3^2 \right]^{-1/2},$$

$$F^{\text{ret}} = \frac{e}{4\pi} \frac{x \wedge \dot{z}}{(\langle \dot{z} | x \rangle^2 + x^2)^{3/2}}$$

$$= \frac{e}{4\pi} \begin{vmatrix} 0 & x_1 - vt & x_2 & x_3 \\ -x_1 + vt & 0 & -vx_2 & -vx_3 \\ -x_2 & vx_2 & 0 & 0 \\ -x_3 & vx_3 & 0 & 0 \end{vmatrix}$$

$$\times \frac{1 - v^2}{[(x_1 - vt)^2 + (1 - v^2)(x_2^2 + x_3^2)]^{3/2}}.$$

Note that

(i) If $v = 0$, the above expression reduces to the usual Coulomb field, $A_0 = e/4\pi r$, $\mathbf{A} = \mathbf{0}$; $\mathbf{E} = e\mathbf{x}/4\pi r^3$, $\mathbf{B} = \mathbf{0}$. When $v$ is not zero, the Coulomb field is simply transformed according to the transformation

law for $E_2(\mathbb{R}^4)$ (cf. (I: 5.2.7)), which is automatically taken into account by the covariant notation.

(ii) The electric field points at the simultaneous position of the charge, not to its retarded position. This fact was used in (1.1.5).

(iii) The denominator contains the spatial distance to $z(s_0)$ in the rest-frame of the charge, viz.,

$$\langle \dot{z} | x - z \rangle = (\langle \dot{z} | x)^2 + x^2)^{1/2} = \left( \frac{(x_1 - vt)^2}{1 - v^2} + x_2^2 + x_3^2 \right)^{1/2}.$$

This is equal neither to the distance $((x_1 - vt)^2 + x_2^2 + x_3^2)^{1/2}$ from the simultaneous position of the charge, nor to the distance from $z(s_0)$, which would be $\langle dt | x - z \rangle$. On the spacelike section $t = 0$, $\langle \dot{z} | x - z \rangle$ equals $r = |\mathbf{x}|$ for the points perpendicular to the direction of motion (i.e., $x_1 = 0$), and is otherwise greater than $r$. The increase of the denominator is compensated for by the factor $1/\sqrt{1 - v^2}$ of the Lorentz transformation, and the net effect is that the static Coulomb potential is altered as shown in Figure 24: $A^0$ is increased perpendicular to the motion and left unchanged in the $x_1$-direction. The oblately squashed potential produces an electric field increased with respect to $\mathbf{x}/4\pi r^3$ if $\mathbf{x} \perp \mathbf{v}$ and decreased if $\mathbf{x} \parallel \mathbf{v}$. The increased range of the Coulomb field makes a charged particle at nearly the speed of light cause greater ionization.

(iv) Of course, the decrease in the field in the forward direction changes into an increase when the charge gets near the reference point. At the time $t = r$, the denominator $\langle \dot{z} | x - z \rangle$ becomes

$$\frac{(x_1 - vr)^2}{1 - v^2} + x_2^2 + x_3^2 = \frac{(r - x_1 v)^2}{1 - v^2}$$

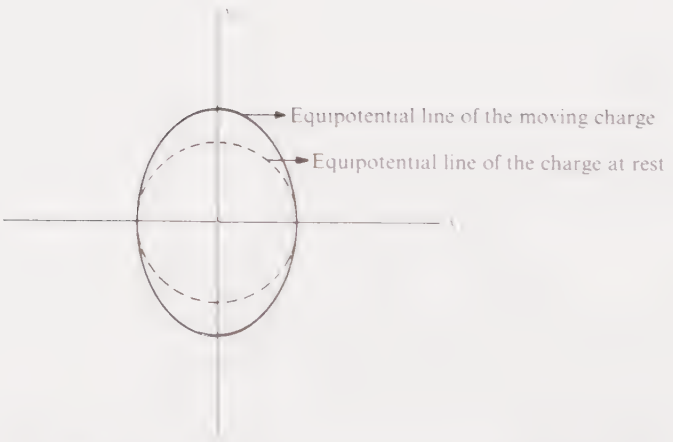



Figure 24 Lorentz contraction of the Coulomb potential

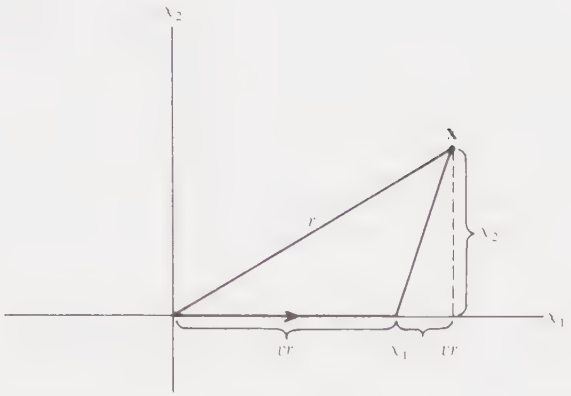



Figure 25 The lengths involved in the field of a point charge

and for $x_2 = x_3 = 0$ is decreased compared with $r^2$ by the factor $(1 - v)/(1 + v)$ (Figure 25). Incidentally, this factor causes the radiation to bunch strongly in the forward direction (Problem 3). The expression $\langle \dot{z} | x - z \rangle$ must be evaluated at the retarded time, and radiation, once emitted, is not affected if the particle flies off afterwards in some different direction, never to come near the reference point.

2. Uniform acceleration. Hyperbolic motion, like that of a charged particle in a constant electric field (cf. I §4.2), is characterized by $z(s)^2 = \text{constant}$. Let

$$z(s) = (\text{Sinh } s, \text{Cosh } s, 0, 0) = \ddot{z}, \dot{z} = (\text{Cosh } s, \text{Sinh } s, 0, 0) = \ddot{z}.$$

Then $s_0$ can be calculated from $(x - z(s_0))^2 = x^2 + 1 - 2\langle x | z \rangle = 0$. Although the equation for $s_0$ is transcendental, it is easy to calculate that

$$\dot{z}(s_0) = \left( \frac{t\xi - x_1(1 + x^2)}{2(t^2 - x_1^2)}, \frac{x_1\xi - t(1 + x^2)}{2(t^2 - x_1^2)}, 0, 0 \right),$$

where

$$\xi \equiv -2\langle \dot{z} | x - z \rangle = [(1 + x^2)^2 + 4(t^2 - x_1^2)]^{1/2}. \qquad (2.3.12)$$

Then

$$A^{\text{ret}} = \frac{-e}{4\pi} \frac{1}{\xi(t^2 - x_1^2)} [t\xi - x_1(1 + x^2), x_1\xi - t(1 + x^2), 0, 0]$$

if $t > -x_1$, and otherwise 0. The general expression (2.3.5) simplifies due to the special properties of the motion $\ddot{z} = z$, $\langle \dot{z} | z \rangle = 0$, and $\langle x | z(s_0) \rangle = (x^2 + 1)/2$:

$$F^{\text{ret}} = \frac{e}{4\pi} \langle \dot{z} | x \rangle^{-2} \left( \dot{z} \frac{x^2 + 1}{2\langle \dot{z} | x \rangle} - z \right) \wedge (x - z).$$

Substitution from (2.3.12) reveals (Problem 1) that, using cylindrical coordinates about the $x_1$-axis, $\rho^2 = x_2^2 + x_3^2$, the only nonzero components are

$$(E_1, E_\rho, B_\varphi) = \frac{e}{4\pi} \left[ \left( \frac{x^2 - 1}{2} \right)^2 + \rho^2 \right]^{-3/2} \left( \frac{x^2 - 1}{2} - \rho^2, \rho x_1, \rho t \right) \quad (2.3.13)$$

if $t > -x_1$, and otherwise these too are zero. Note that

(i) The part of space where $t + x_1 \leq 0$ has no field, because it can not be connected to the world-line by any light-cone (Figure 26):

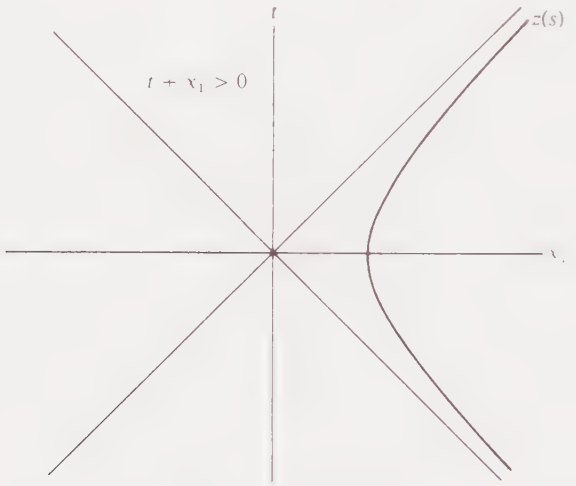



Figure 26 Hyperbolic motion

(ii) As a practical matter, motion is never strictly hyperbolic. If hyperbolic motion is combined with uniform motion at a velocity $-v$ (respectively $v$) at the point $(-v, 1, 0, 0)/\sqrt{1 - v^2}$ (resp. $(v, 1, 0, 0)/\sqrt{1 - v^2}$), then the results of Examples 1 and 2 in the appropriate regions are simply combined (see Figure 27). Interestingly, this field does not converge to (2.3.13) as $v \to 1$, as there remains a contribution from the initial uniform motion:

$$\lim_{v \to 1} \frac{e}{4\pi} \frac{(x_1 - \sqrt{1 - v^2} + vt, \rho, -v\rho)(1 - v^2)}{[(x_1 - d + vt)^2 + (1 - v^2)\rho^2]^{3/2}}$$

$$= \frac{e}{2\pi} \frac{\delta(x_1 + t)}{1 + \rho^2} (0, \rho, -\rho)$$

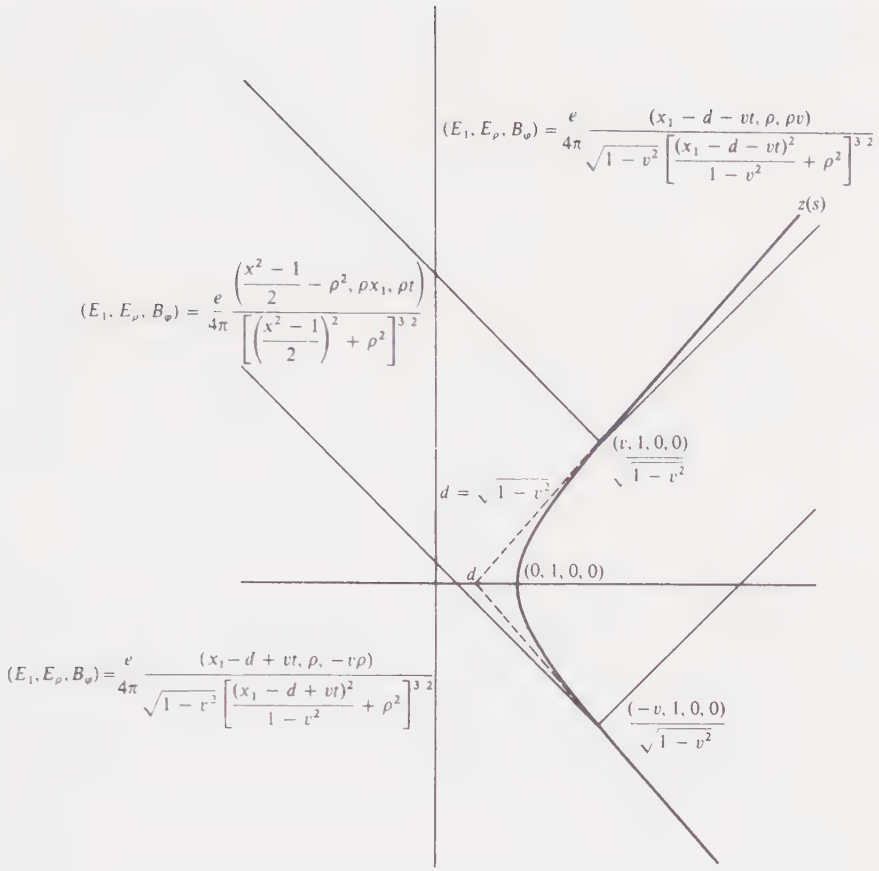

$(E_1, E_\rho, B_\varphi) = \frac{e}{4\pi} \frac{(x_1 - d - vt, \rho, \rho v)}{\sqrt{1-v^2}\left[\frac{(x_1 - d - vt)^2}{1-v^2} + \rho^2\right]^{3/2}}$
$z(s)$
$(E_1, E_\rho, B_\varphi) = \frac{e}{4\pi} \frac{\left(\frac{x^2-1}{2} - \rho^2, \rho x_1, \rho t\right)}{\left[\left(\frac{x^2-1}{2}\right)^2 + \rho^2\right]^{3/2}}$
$(v, 1, 0, 0)$
$\sqrt{1-v^2}$
$d = \sqrt{1-v^2}$
$d$
$(0, 1, 0, 0)$
$(E_1, E_\rho, B_\varphi) = \frac{e}{4\pi} \frac{(x_1 - d + vt, \rho, -v\rho)}{\sqrt{1-v^2}\left[\frac{(x_1 - d + vt)^2}{1-v^2} + \rho^2\right]^{3/2}}$
$(-v, 1, 0, 0)$
$\sqrt{1-v^2}$

Figure 27 Hyperbolic motion followed by uniform motion

(see Problem 5). This field, which has accumulated on the surface $x_1 = -t$, must be added to (2.3.13), and it also shows up in other physically meaningful limiting processes, such as when the hyperbolic motion is that of one of the particles of a pair-production [38].

(iii) Equation (2.3.13) is not acceptable mathematically, because

$$\delta F^{\text{ret}} - J = \frac{e}{\pi} \frac{\delta(x_1 + t)}{(1 + \rho^2)^2} (1, 1, 0, 0).$$

The apparent surface current vanishes if the field of (ii) is added to $F^{\text{ret}}$. Taken by itself, $F^{\text{ret}}$ is neither the limit of the $F$ of (ii) as $v \to 1$, nor the limit as $T \to \infty$ of an $F_T^{\text{ret}}$ that would result from using the half-space $t \geq -T$ for $N$ in (2.2.9). As noted in (2.2.10; 2), $\delta F_T^{\text{ret}} - J_{|N} = -dj$, but the right side of this equation approaches zero as $T \to \infty$

(Problem 6), while $\delta F^{\text{ret}} - J \sim \delta(x_1 + t)$. Its integral over $x_1$ does not tend to zero, although the integral of $dj$ does.

(iv) If we write the field of a uniformly moving charge in the notation of (2.3.13),

$$(E_1, E_\rho, B_\varphi) = \frac{e}{4\pi}\left[\frac{(x_1 - vt)^2}{1 - v^2} + \rho^2\right]^{-3/2} \times \left(\frac{x_1 - vt}{\sqrt{1 - v^2}}, \frac{\rho}{\sqrt{1 - v^2}}, \frac{v\rho}{\sqrt{1 - v^2}}\right),$$

then we see that the term $(x_1 - vt)/\sqrt{1 - v^2}$ is replaced in (2.3.13) with $(x^2 - 1)/2$ and that $E_1$ contains an additional term $-\rho^2[\ldots]^{-3/2}$. This is the compression factor of the intuitive picture (1.1.5).

(v) At the time when $x^2 = 1$, the fields become

$$(E_1, E_\rho, B_\varphi) = \frac{e}{4\pi\rho}\left(-1, \frac{x_1}{\rho}, \sqrt{1 + (1 + x_1^2)/\rho^2}\right).$$

The maximum field strength over time consequently falls off only as $1/\rho$ as a function of the distance $\rho$ to the line of flight of the charge.

(vi) Except for the condition $t + x_1 > 0$, $E_1$ and $B_\varphi$ are even in $x_1$, while $E_\rho$ is odd. The fields are just as large where $x_1 < 0$, where the charge never goes, as where $x_1 > 0$. The total radiation field over the line $t = -x_1$ is therefore just as large as the total Coulomb field over the line $t = x_1$, which the world-line of the particle asymptotically approaches.

(vii) Again disregarding the condition that $t + x_1 > 0$, $\mathbf{E}$ is even in $t$ and $\mathbf{B}$ is odd. In particular, the magnetic field is zero throughout space at the time $t = 0$.

3. Rotating charges. If

$$z(s) = \left(\frac{s}{\sqrt{1 - v^2}}, R\cos\frac{vs/R}{\sqrt{1 - v^2}}, R\sin\frac{vs/R}{\sqrt{1 - v^2}}, 0\right), \quad (2.3.14)$$

the determination of $s_0(x)$ is more difficult than in Example 2, and therefore one usually looks only at the limit $v \to 0$, $R \to 0$, $e \to \infty$, such that $v/R \to \omega$ and $e = 1/R\omega^2$. In order not to be encumbered with the infinite Coulomb field that results, one considers two opposite charges in mirror-image paths about the origin (Hertz's dipole). In this limit, $s_0 = t - r$, $z - x = (r, \mathbf{x})$, $\dot{z} = (1, 0, 0, 0)$, $e\ddot{z} = -(0, \cos\omega(r - t), \sin\omega(r - t), 0)$, and the fields become

$$\mathbf{B} = \frac{e[\ddot{\mathbf{z}} \times \mathbf{x}]}{4\pi r^2}, \qquad \mathbf{E} = \frac{[\mathbf{B} \times \mathbf{x}]}{r}, \qquad (2.3.15)$$

$$e\ddot{\mathbf{z}} = (\cos\omega(r - t), \quad \sin\omega(r - t), 0).$$

These examples illustrate $F$ in three representative cases, for free motion and linear and circular acceleration. Often of greater practical interest than the field strengths are the energy and momentum forms created by the charge. These will be sums (1.3.22) of two terms quadratic in $F$. Since each component of $F$ is itself a sum of six fractions, it seems that blind substitution would produce 72 fractions. Fortunately, the special structure of the $F$ of a point charge can be used to reduce the algebraic complexity.

**The Energy-Momentum Forms of the Field of a Point Charge** (2.3.16)

With the rules (1.2.18), $\mathscr{T}_\alpha$ can be rewritten

$$\mathscr{T}_\alpha = *((i_\alpha F) \wedge *F - \tfrac{1}{2} i_\alpha (F \wedge *F)) = \frac{e_\alpha}{2} *(F \wedge *F) - i_{i_\alpha F} F.$$

Our $F$ is of the form

$$F = \frac{e}{4\pi} \langle \dot{z} | x - z \rangle^{-2} v \wedge n, \qquad v = \dot{z} \frac{1 + \langle \ddot{z} | x - z \rangle}{\langle \dot{z} | x - z \rangle} - \ddot{z}, \qquad n = (x - z),$$

and so the invariants are

$$n^2 = 0, \qquad \langle n | v \rangle = 1, \qquad v^2 = \ddot{z}^2 - \left( \frac{1 + \langle \ddot{z} | x - z \rangle}{\langle \dot{z} | x - z \rangle} \right)^2$$

(use $\dot{z}^2 = -1$ and $\langle \dot{z} | z \rangle = 0$). To calculate $\mathscr{T}_\alpha$ we need the equations

$$*(v \wedge n \wedge *(v \wedge n)) = i_v i_n v \wedge n = 1, \; i_\alpha v \wedge n = v_\alpha n - n_\alpha v,$$

and

$$i_{v_\alpha n - n_\alpha v} v \wedge n = v_\alpha n \langle n | v \rangle - n_\alpha (n v^2 - v \langle n | v \rangle).$$

By substituting for the scalar product, one finds that

$$\mathscr{T}_\alpha = \left( \frac{e}{4\pi} \right)^2 \langle \dot{z} | x - z \rangle^{-4} \Bigg\{ (x - z)(x - z)_\alpha \left( \ddot{z}^2 - \left( \frac{1 + \langle \ddot{z} | x - z \rangle}{\langle \dot{z} | x - z \rangle} \right)^2 \right)$$

$$+ \ddot{z}(x - z)_\alpha + (x - z)\ddot{z}_\alpha - (\dot{z}(x - z)_\alpha + (x - z)\dot{z}_\alpha)$$

$$\times \frac{1 + \langle \ddot{z} | x - z \rangle}{\langle \dot{z} | x - z \rangle} + \frac{e_\alpha}{2} \Bigg\}.$$

**Remarks** (2.3.17)

1. We have only used $F^{\text{ret}}$, which corresponds to the initial condition $F^{\text{in}} = 0$.
2. The terms that contain $\ddot{z}$ quadratically are recognizable as the $\mathscr{T}_\alpha$ of $F^{(\ddot{z})}$. At large distances they would dominate, as they decrease as $1/r^2$. The contribution from $F^{(\dot{z})}$ goes as $1/r^4$, and the mixed term as $1/r^3$.

3. The structure of $\mathcal{T}_\alpha$ shows $T_{\alpha\beta} = T_{\beta\alpha}$ and $T_\alpha^\alpha = 0$.

Poynting's vector $\mathcal{S}_j = [\mathbf{E} \times \mathbf{B}]_j$ is useful for visualizing how the field energy flows. This can best be understood by returning to the representative

**Examples** (2.3.18)

1. Uniform motion. In cylindrical coordinates, $F$ of the form given in (2.3.11; 2), paragraph (iv), makes Poynting's vector

$$(\mathcal{S}_1, \mathcal{S}_\rho, \mathcal{S}_\varphi) = \left(\frac{e}{4\pi}\right)^2 \frac{\rho v(1 - v^2)}{[(x_1 - vt)^2 + (1 - v^2)\rho^2]^3}(\rho, -x_1 + vt, 0).$$

The streamlines of energy are circles $\rho^2 + (x - vt)^2 = R^2$ around the (simultaneous, not retarded) position of the charge $vt$. The field energy flows toward the future positions of the charge, in other words, along with the charge. (See Figure 28.)

2. Uniform acceleration. From (2.3.11; 2),

$$(\mathcal{S}_1, \mathcal{S}_\rho, \mathcal{S}_\varphi) = \left(\frac{e}{4\pi}\right)^2 \frac{\rho t}{\left[\left(\frac{x^2 - 1}{2}\right)^2 + \rho^2\right]^3}\left(x_1\rho, \rho^2 + \frac{1 - x^2}{2}, 0\right),$$

in the notation used above. The streamlines are again circles, $\rho^2 + (x_1 - \sqrt{R^2 + t^2 + 1})^2 = R^2$, but as the radius $R$ increases, the center of the circle moves ahead of the position of the charge. This makes the flow of energy point more and more outward as $\rho$ increases with fixed $x_1$ (Figure 28); this occurs because $\mathbf{E}$ has a stronger component in the direction of the motion.

3. A rotating charge in the dipole limit. With (2.3.15), we get

$$\mathcal{S} = \frac{\mathbf{x}}{(4\pi)^2 r^3}\left(|\ddot{\mathbf{z}}|^2 - \left(\ddot{\mathbf{z}} \cdot \frac{\mathbf{x}}{r}\right)^2\right) = \frac{\mathbf{x}}{(4\pi)^2 r^3}(1 - \sin^2\vartheta \cos^2(\omega(r - t) - \varphi))$$

in polar coordinates. The energy flows radially outward, and the flow is strongest perpendicular to the direction of the acceleration at the retarded time.

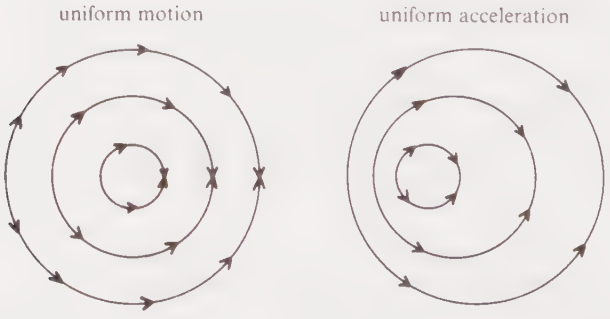



Figure 28 Streamlines of the energy

**Remarks** (2.3.19)

1. If we return to the case of uniform motion at the velocity $\pm v$ joined to hyperbolic motion, then we have to add the streamlines of Examples 1 and 2. In the spherical shell

$$\left(t - \frac{v}{\sqrt{1-v^2}}\right)^2 < \left(x - \frac{1}{\sqrt{1-v^2}}\right)^2 + \rho^2 < \left(t + \frac{v}{\sqrt{1-v^2}}\right)^2$$

we would obtain an increased flow of energy outwards.

2. Opinions differ as to whether a charge in hyperbolic motion emits radiation, owing to the different possible definitions made when posing global questions like that of radiation to infinity. In §3.4 a local definition will be introduced, using the reaction of the radiation on the charge. At this point we summarize the facts supporting the various opinions, and leave the reader to make up his own mind:

   (a) $\mathrm{Max}_t |\mathscr{S}|$ falls off only as $1/\rho^2$.
   (b) Expressions like $\int_K d\mathbf{S} \cdot \mathscr{S}(t = R)$, where $K$ is a sphere of radius $R$ centered on the position of the particle, tend to positive values as $R \to \infty$.
   (c) All the streamlines are closed, and none of them run to infinity.

Since the field equations are linear, the spectral distribution of the sources is inherited by the field. We therefore end this section with a brief discussion of the

**Fourier Decomposition of the Current of a Point Charge** (2.3.20)

$$\tilde{J}_\alpha(k) = \int d^4x\, e^{-i\langle k|x\rangle} J_\alpha(x) = e \int_{-\infty}^{\infty} ds\, \dot{z}_\alpha(s) e^{-i\langle k|z(s)\rangle}.$$

**Examples** (2.3.21)

1. The sudden acceleration of a charge,

$$z(s) = \begin{cases} \dot{\bar{z}} s & \text{if} \quad s > 0, \qquad \dot{\bar{z}} = \dfrac{1}{\sqrt{1-\bar{v}^2}}(1, \bar{\mathbf{v}}) \\ \dot{z} s & \text{if} \quad s \le 0, \qquad \dot{z} = \dfrac{1}{\sqrt{1-v^2}}(1, \mathbf{v}). \end{cases}$$

The distribution $\tilde{J}_\alpha(k)$ turns out to be

$$\tilde{J}_\alpha(k) = ie \lim_{\varepsilon \to 0} \left( \frac{\dot{z}_\alpha}{\langle k|\dot{z}\rangle + i\varepsilon} - \frac{\dot{\bar{z}}_\alpha}{\langle k|\dot{\bar{z}}\rangle - i\varepsilon} \right).$$

Consequently,

(i) if $\dot{z} = \bar{z}$, then $\tilde{J}_\alpha(k) = e2\pi\dot{z}_\alpha(\langle k|\dot{z}\rangle)$, and if $v = \bar{v} = 0$, $\tilde{J}_\alpha(k)$ is thus $\sim\delta(k^0)$.

(ii) The use of this $\tilde{J}(k)$ in (2.2.23), although it does not satisfy the assumptions needed to justify that formula, produces

$$P^\alpha(\infty) - P^\alpha(-\infty) = e^2 \int \frac{d^4k}{(2\pi)^3}\,\delta(k^2)k^\alpha\Theta(k^0)\left|\frac{\dot{z}}{\langle\dot{z}|k\rangle} - \frac{\dot{\bar{z}}}{\langle\dot{\bar{z}}|k\rangle}\right|^2.$$

It might at first be thought that (2.2.19) could not be used for a point charge, since the field at the position of the particle, and hence also $F^{\text{ret}} \wedge {}^*J$, is infinite. As we shall learn in §2.4, however, $F^{\text{rad}} \wedge {}^*J$ remains finite, and so under the right circumstances (2.2.23) is also applicable to point particles. It should furthermore be no surprise that the integral diverges for large $|\mathbf{k}|$, as $\ddot{z}$, and consequently also $F^{(z)}$, become infinite for this motion. It can be hoped that if the acceleration were made gentler during a time $\tau$, the integrand could be suppressed for $|\mathbf{k}| > 1/\tau$, allowing the integral to converge, while not essentially changing $\tilde{J}$ for $|\mathbf{k}| < 1/\tau$. With this motive, let us leave the integral as it is, but rewrite it by using the continuity equation for current,

$$k^\alpha\tilde{J}_\alpha = k^0\left(\tilde{J}^0 - \left(\frac{\mathbf{k}}{k^0}\cdot\tilde{\mathbf{J}}\right)\right) = 0.$$

This brings the spatial components $\tilde{\mathbf{J}}_\perp$ of $\tilde{J}$ that are perpendicular to $\mathbf{k}$ into play: If $k^2 = 0$, then

$$|\tilde{J}^\alpha\tilde{J}_\alpha| = |\tilde{J}_\perp|^2, \qquad \tilde{\mathbf{J}}_\perp \equiv \tilde{\mathbf{J}} - \frac{\mathbf{k}(\mathbf{k}\cdot\tilde{\mathbf{J}})}{|\mathbf{k}|^2}.$$

Defining $\mathbf{w} \equiv \mathbf{v}/(1 - |\mathbf{v}|\cos\vartheta)$, $\vartheta = \sphericalangle(\mathbf{k}, \mathbf{v})$, and defining $\bar{\mathbf{w}}$ analogously, makes the energy loss

$$P^0(\infty) - P^0(-\infty) = e^2 \int \frac{d^3k}{(2\pi)^3}\,\frac{|\mathbf{w} - \bar{\mathbf{w}}|_\perp^2}{2|\mathbf{k}|^2}.$$

If $|\mathbf{v}|$ and $|\bar{\mathbf{v}}| \ll 1$, then $|\mathbf{w} - \bar{\mathbf{w}}|_\perp^2 = |\mathbf{v} - \bar{\mathbf{v}}|^2 - \langle\mathbf{k}|\mathbf{v} - \bar{\mathbf{v}}\rangle^2/|\mathbf{k}|^2$, and thus the maximum occurs for $\mathbf{k}\perp\mathbf{v} - \bar{\mathbf{v}}$; for wave-vectors perpendicular to the acceleration there is enhancement. For relativistic motion, the denominators $1 - v\cos\vartheta$ and $1 - \bar{v}\cos\vartheta$ strongly favor the directions of $\mathbf{v}$ and $\bar{\mathbf{v}}$ (see Problem 3). The frequency distribution has the characteristic spectrum $d^3k/|\mathbf{k}|^2 \sim dk$ of bremsstrahlung; the same energy is radiated in every frequency interval.

2. A rotating charge. Whereas the field of a Hertz dipole (2.3.15) is purely harmonic in time, the fields of charges rotating in circles with finite radii also exhibit higher harmonic frequencies. Thus it comes about that highly energetic electrons in a magnetic field emit a characteristic x-ray spectrum. Let us calculate the energy emitted at a given harmonic frequency.

A formula analogous to (2.2.23) can be derived (Problem 7) for the energy lost per period $\omega$ by a current with periodic time-dependence. The only essential change is the definition of $\tilde{J}$, as we now use a Fourier series in time:

$$\tilde{J}^\alpha(n\omega, \mathbf{k}) \equiv \frac{\omega}{2\pi} \int_0^{2\pi/\omega} dt \int_{\mathbb{R}^3} d^3x\, e^{-i(\mathbf{k}\cdot\mathbf{x} - n\omega t)} J^\alpha(x), \qquad n \in \mathbb{Z}. \quad (2.3.22)$$

There results

$$P^0\left(\frac{2\pi}{\omega}\right) - P^0(0) = \sum_{n \geq 1} \int \frac{d^3k}{(2\pi)^3} 2\pi\delta(|\mathbf{k}|^2 - (n\omega)^2) n\omega \cdot \tilde{J}^\beta(n\omega, \mathbf{k}) \tilde{J}_\beta(-n\omega, -\mathbf{k}). \quad (2.3.23)$$

If the current (2.3.14) is written as

$$J^\alpha(x) = e \int_{-\infty}^{\infty} ds\, \dot{z}^\alpha(s) \delta^4(x - z(s)) = e(1, -v \sin \omega t, v \cos \omega t, 0) \delta^3(\mathbf{x} - \mathbf{z}(t)),$$

then

$$\tilde{J}(n\omega, \mathbf{k}) = e\omega \int_0^{2\pi/\omega} \frac{dt}{2\pi} e^{i(n\omega t - \mathbf{k}\cdot\mathbf{z}(t))} (1, -v \sin \omega t, v \cos \omega t, 0). \quad (2.3.24)$$

Since the radiation is concentrated in the plane of motion within a sector of opening angle $\sim \sqrt{1 - v}$ (Problem 3), it is interesting to calculate the frequency distribution for $\mathbf{k}$ in the plane of motion, let us say in the $x_2$-direction. According to the argument in the preceding example, only $J^1$ contributes to the energy loss. The integral (2.3.24) becomes a Bessel function

$$\tilde{J}^1(n\omega, 0, n\omega, 0) = -e\omega v \int_0^{2\pi/\omega} \frac{dt}{2\pi} \sin \omega t e^{in(\omega t - v \sin \omega t)} = -ieR J_n'(nv). \quad (2.3.25)$$

To discuss this result, it is necessary to distinguish the cases $nv \ll 1$ and $nv \gg 1$. In the former case, one can use the well-known expansion ([20], cf. (3.4.3) and (3.4.4)):

$$J_n'(nv) = \frac{1}{2(n-1)!} \left(\frac{nv}{2}\right)^{n-1} (1 + O((nv)^2)). \quad (2.3.26)$$

This shows that the frequency distribution of (2.3.23) is $\sim (nv)^{2n}/[(n-1)!]^2$, which has a strong maximum at $n = 1$ if $v \ll 1$. If $v = 1$, the simple formula

$$J_n'(n) \to \frac{2^{2/3}}{3^{1/3}\Gamma(\frac{1}{3})} n^{-2/3}$$

(see [20], 9.4.43) could be used, and the spectrum would be $n^2 |\tilde{J}|^2 \sim n^{2/3}$. Since $v < 1$, however, this formula is not valid for high $n$, and the better

asymptotic formula,

$$J_n'(nv) \to -\frac{2}{v}\left(\frac{1-v^2}{4\xi}\right)^{1/4}\frac{\mathrm{Ai}'(n^{2/3}\xi)}{n^{2/3}}$$

$$\tfrac{2}{3}\xi^{3/2} = \ln\frac{1+\sqrt{1-v^2}}{v} - \sqrt{1-v^2} \underset{v\to 1}{\to} \frac{(1-v^2)^{3/2}}{3}$$

$$\mathrm{Ai}'(z) \begin{cases} \to -\dfrac{1}{3^{1/3}\Gamma(\frac{1}{3})} & \text{as } z \to 0 \\ \to \dfrac{1}{2\sqrt{\pi}} z^{1/4} e^{-(2/3)z^{3/2}} & \text{as } z \to \infty \end{cases} \qquad (2.3.27)$$

must be resorted to. This modifies the spectrum to

$$n \exp\left(-\tfrac{2}{3}n(1-v^2)^{3/2}\right)$$

for $n^{2/3}(1-v^2) \gg 1$. Therefore the spectrum has its maximum at $n \sim (1-v^2)^{-3/2} \equiv \gamma^3$, after which it decreases exponentially.

**Remark** (2.3.28)

The reason that the maximum is at $\sim\gamma^3$ can be understood as follows: An observer $O$ first sees the light that is emitted in an angular interval $\sim\gamma^{-1}$. (See Figure 29.) Since the beam of radiation sweeps just barely ahead of the charge, the light arrives at $O$ within a time interval

$$\Delta t = (1-v)\gamma^{-1}R \sim \gamma^{-3}R$$

(if $v \to 1$), and hence

$$\omega_{\max} \sim \frac{1}{\Delta t} \sim \frac{\gamma^3}{R} \sim \omega\gamma^3.$$

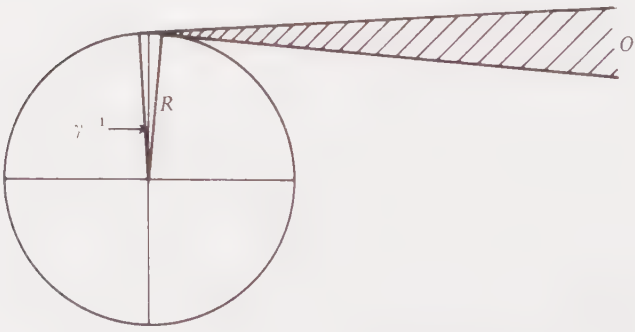



Figure 29 Radiation from relativistic motion in a circle

## Problems (2.3.29)

1. Substitute from (2.3.12) into the expression for $F^{\text{ret}}$.

2. Calculate the field and Poynting's vector $\mathscr{S}$ for a charge moving in a straight line at the speed of light.

3. Calculate the energy distribution $T_{00}$ of the radiation of a fast-moving charge $[\dot{z}(s_0) = 1/\sqrt{1 - v^2}\,(1, 0, 0, v), z(s_0) = 0]$ in the far zone at $x^2 = 0$ and

   (a) longitudinal acceleration ($\ddot{z}(s_0) \sim (v, 0, 0, 1)$); and
   (b) transverse acceleration ($\ddot{z}(s_0) \sim (0, 0, 1, 0)$).

   At what angle, in the usual polar coordinates, is the maximum as $v \to 1$?

4. Calculate $\int_{t=0} \mathscr{T}^z$ for the field of a uniformly moving point charge (2.3.11; 1). Use $\bar{x} = (x_1/\sqrt{1 - v^2},\ x_2,\ x_3)$ as the integration variables, and define the classical radius of the electron by the divergent integral

$$\frac{1}{r_c} = \frac{1}{2}\int \frac{d^3\bar{x}}{|\bar{\mathbf{x}}|^4}.$$

   Show that this destroys the vector transformation property of (1.3.37; 5). How come?

5. Show that

$$\lim_{v \to 1} \frac{ed^2}{4\pi} \frac{(x_1 + vt - d, \rho, -v\rho)\Theta((\rho^2 + 1)d/2 - x_1 - vt)}{[(x_1 + vt - d)^2 + \rho^2 d^2]^{3/2}}$$

$$= \frac{e}{4\pi} \frac{2\rho}{1 + \rho^2} \delta(t + x_1)(0, 1, -1),$$

   where $d \equiv \sqrt{1 - v^2}$.

6. Calculate $j(x) = \int_{t=-T} d^3x\, D^{\text{ret}}(\bar{x} - x)J_0(x)$ from (2.2.10; 2) for a point charge, and take the limit as $T \to \infty$ of $j$ for hyperbolic motion.

7. Calculate $P^0(2\pi\,\omega) - P^0(0)$ for a $J$ that depends on time periodically, by using the Fourier series (2.3.22).

## Solutions (2.3.30)

1. $((x^2 + 1)/2\langle\dot{z}|x\rangle)\dot{z} - z = -(2/\xi)(x_1, t, 0, 0)$.

$$F^{\text{ret}} = \frac{-8}{\xi^3}(x_1, t, 0, 0)$$

$$\wedge \left(\frac{t(t^2 - x_1^2 + \rho^2 + 1) - x_1\xi}{2(t^2 - x_1^2)}, \frac{x_1(t^2 - x_1^2 + \rho^2 + 1) - t\xi}{2(t^2 - x_1^2)}, x_2, x_3\right) \frac{e}{4\pi}$$

   Using $\mathbf{x} \cdot d\mathbf{x} = x_1\, dx_1 + \rho\, d\rho$, we find (2.3.13).

2. The limit as $v \to 1$ of the fields of (2.3.11; 1) is

$$(E_1, E_\rho, E_\varphi) = \frac{e}{2\pi}\delta(x - t)\frac{1}{\rho}(0, 1, 1);$$

and this satisfies $\delta F = J$ with

$$J = e\delta(x - t)\delta(y)\delta(z)(1, 1, 0, 0) = \lim_{v \to 1} e\delta(x - vt)\delta(y)\delta(z)(1, v, 0, 0).$$

$\mathscr{S}$ points in the direction $x_1$, but because of the factor $\delta^2$, this infinitely Lorentz-contracted field has an infinite energy-momentum density.

3. (a) $T_{00} \sim \dfrac{(1 - v^2)^4 \sin^2 \vartheta}{r^2(1 - v\cos\vartheta)^6}, \qquad \vartheta_{\max} \sim \sqrt{\dfrac{1}{v} - 1} \sim \dfrac{m}{E}.$

(b) $T_{00} \sim \dfrac{(1 - v^2)^2}{r^2(1 - v\cos\vartheta)^6}[(1 - v\cos\vartheta)^2 - (1 - v^2)\sin^2\vartheta\sin^2\varphi].$

4. $$\int \mathscr{T}^0 = \frac{1 + v^3/3}{1 - v^2}\frac{1}{r_c}, \qquad \int \mathscr{T}^j = v^j \frac{4/3}{1 - v^2}\frac{1}{r_c} \neq v^j \int \mathscr{T}^0(v = 0)/\sqrt{1 - v^2}.$$

The condition that $\delta\mathscr{T}^\alpha = 0$, used in the derivation of the transformation law, is violated at the origin. Even adding a $t^\alpha$ for the particle will not avoid the problem, as $P^\alpha$ equals $(m/\sqrt{1 - v^2}, m\mathbf{v}/\sqrt{1 - v^2})$. Thus $\delta(\mathscr{T}^\alpha + t^\alpha)$ is only formally zero (cf. (1.3.25; 2) and (2.1.18; 7)).

5. If $x_1 \neq t$, then the three components approach zero; on the other hand, with $\alpha = (x + vt)/d - 1$,

$$\begin{aligned}\int_{-\infty}^{\infty} \frac{dt\, d^2\Theta(\ )}{[(x_1 + vt - d)^2 + \rho^2 d^2]^{3/2}} &\to \int_{-\infty}^{d(1+\rho^2)/2 - x_1} \frac{dt\, d^2}{[(x_1 + vt - d)^2 + \rho^2 d^2]^{3/2}} \\ &\to \frac{1}{v}\int_{-\infty}^{(\rho^2-1)/2} \frac{d\alpha}{[\alpha^2 + \rho^2]^{3/2}} \\ &= \frac{1}{v\rho^2}\frac{\alpha}{(\alpha^2 + \rho^2)^{1/2}}\Bigg|_{-\infty}^{(\rho^2-1)/2} \\ &= \frac{1}{v\rho^2}\left(1 + \frac{\rho^2 - 1}{\rho^2 + 1}\right) \to \frac{2}{1 + \rho^2}.\end{aligned}$$

6. If $z$ is the point at which the world-line $z(s)$ crosses the hypersurface at $t = -T$, then

$$j(x) = \frac{e}{2\pi}\delta((x - z)^2)\Theta(x^0 - z^0).$$

If the motion is hyperbolic,

$$z = \left(\frac{-v}{\sqrt{1 - v^2}}, \frac{1}{\sqrt{1 - v^2}}, 0, 0\right),$$

and as $v \to 1$,

$$\delta(x^2 + 1 - 2\langle x|z\rangle) = \sqrt{1 - v^2}\,\delta(2(x + vt) - (x^2 + 1)\sqrt{1 - v^2}) \to 0.$$

7. Substituting the Fourier series

$$J(x) = \sum_{n \in \mathbb{Z}} \int \frac{d^3k}{(2\pi)^3} \tilde{J}(n\omega, \mathbf{k}) e^{i(\mathbf{k}\cdot\mathbf{x} - n\omega t)}$$

into the integral

$$\begin{aligned} P^0(2\pi/\omega) - P^0(0) &= \int_{0 \le \bar{t} \le 2\pi/\omega} i_0 F^{\text{ret}} \wedge {}^*J \\ &= \int_{0 \le \bar{t} \le 2\pi/\omega} d^4\bar{x} J^\beta(\bar{x}) \int_{\mathbb{R}^4} d^4x \left( J^0(x) \frac{\partial}{\partial \bar{x}^\beta} D^{\text{ret}}(\bar{x} - x) \right. \\ &\quad \left. - J_\beta(x) \frac{\partial}{\partial \bar{x}^0} D^{\text{ret}}(\bar{x} - x) \right) \end{aligned}$$

and using (2.2.4) and (2.2.5) yields

$$\int d^4x \, D^{\text{ret}}(\bar{x} - x) J^\alpha(x) = \sum_{n \in \mathbb{Z}} \int \frac{d^3k}{(2\pi)^3} \tilde{J}(n\omega, \mathbf{k}) \frac{e^{i(\mathbf{k}\cdot\bar{\mathbf{x}} - n\omega t)}}{|\mathbf{k}|^2 - (n\omega + i\varepsilon)^2}.$$

Then, finally,

$$P^0\left(\frac{2\pi}{\omega}\right) - P^0(0) = \sum_{n \ge 1} \int \frac{d^3k}{(2\pi)^3} 2\pi\delta(|\mathbf{k}|^2 - (n\omega)^2) n\omega \tilde{J}^\beta(n\omega, \mathbf{k}) \tilde{J}_\beta(-n\omega, -\mathbf{k}).$$

## 2.4 Radiative Reaction

*The radiation of electromagnetic energy causes a reactive force on a charge. The calculation of this force for point particles is tricky, as it involves divergent integrals.*

The product of $J$ and $F$ appears in the Lorentz force (1.3.24), but it is not well defined for a point particle, since the field is singular at the particle's position. Let us first look into the less problematical matter of the energy-momentum of the emitted radiation, as we slowly work up to the infinities in the equation for the total energy and momentum.

The starting point is Stokes's theorem for the electromagnetic energy-momentum forms:

$$\int_N d^*\mathcal{T}^\alpha = \int_{\partial N} {}^*\mathcal{T}^\alpha. \tag{2.4.1}$$

We choose $N$ as the four-dimensional region bounded by the light-cones

$$\begin{aligned} L_1 &= \{x \in \mathbb{R}^4 : (x - z(s_1))^2 = 0, x^0 > z^0(s_1)\} \\ L_2 &= \{x \in \mathbb{R}^4 : (x - z(s_2))^2 = 0, x^0 > z^0(s_2)\} \end{aligned}$$

and the cylinder

$$K = \{x \in \mathbb{R}^4 : (x - z(s_1))^2 + \langle \dot{z}(s_1) | x - z(s_1) \rangle^2 = R^2\},$$

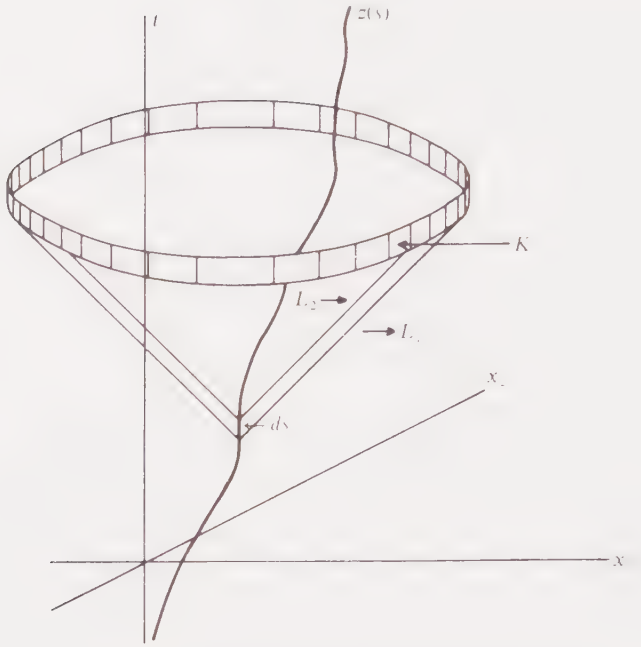



Figure 30 The hypersurface used for $\partial N$ in (2.4.1)

in order to be able to follow the radiation as it goes to infinity (Figure 30). We first calculate the part of (2.4.1) coming from the piece of $\partial N$ contained in $K$. For this purpose $R$ may be increased to $\infty$, and $ds \equiv s_2 - s_1 \to 0$; the result has the interpretation of the amount of energy-momentum that is lost by the charge between $s_1$ and $s_1 + ds$ and escapes to infinity. In this limit $\int_{\partial N \cap K} {}^*\mathscr{T}_\alpha$ consists only of the contribution of $F^{(\dot{z})}$ to $\mathscr{T}_\alpha$ (cf. (2.3.16))

$$\mathscr{T}_\alpha^{(\dot{z})} = \left(\frac{e}{4\pi}\right)^2 \langle \dot{z} | x - z \rangle^{-4} (x - z)(x - z)_\alpha \left\{ \ddot{z}^2 - \left( \frac{\langle \ddot{z} | x - z \rangle}{\langle \dot{z} | x - z \rangle} \right)^2 \right\};$$

the integral of this is asymptotically independent of $R$, while the other terms all have higher powers of $R$ in the denominator. The external surface $\partial N \cap K$ has a height $ds$ above $K \cap L_1$, and in the limit as $ds \to 0$, we need to know ${}^*\mathscr{T}_\alpha$ only on $K \cap L_1$. If we write $x - z = R(\dot{z} + n)$, $n \in E_1$, on that surface (see Figure 31), then from $x \in K \cap L_1$:

$$0 = R^{-2}(x - z)^2 = -1 + 2\langle \dot{z} | n \rangle + n^2$$

and

$$1 = R^{-2}\{(x - z)^2 + \langle \dot{z} | x - z \rangle^2\} = (-1 + \langle z | n \rangle)^2$$

it follows that $n^2 = 1$ and $\langle \dot{z} | n \rangle = 0$, and hence that

$$-\langle \dot{z} | x - z \rangle = R \quad \text{and} \quad \langle \ddot{z} | x - z \rangle = R\langle \ddot{z} | n \rangle.$$

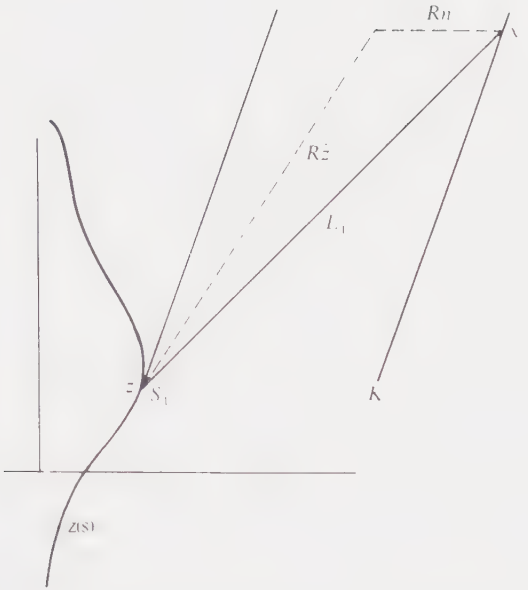



Figure 31 The quantities introduced for the evaluation of (2.4.1)

Consequently,

$$\mathcal{T}_x^{(\dot{z})} = \left(\frac{e}{4\pi}\right)^2 (\dot{z} + n)(\dot{z}_\alpha + n_\alpha)(\ddot{z}^2 - \langle \ddot{z}|n\rangle^2)R^{-2},$$

where $n$ varies over the spatial unit sphere in $K \cap L_{10}$. The 3-form $*(\dot{z} + n)$ acts in the integral $\int *\mathcal{T}_x$ as $R^2\, ds\, d\Omega_n$ ($d\Omega$ is the element of solid angle on the unit sphere). By symmetry, all odd powers of $n$, and hence the part containing $n_x$, drop out of the integral $\int *\mathcal{T}_x$. By taking the average over the unit sphere we simply replace $\langle \ddot{z}|n\rangle^2$ with $\ddot{z}^2/3$, and in this limit we obtain

**Larmor's Formula** (2.4.2)

$$\int_{\partial N \cap K} *\mathcal{T}_x = \frac{2}{3}\frac{e^2}{4\pi} ds\, \dot{z}_x \ddot{z}^2.$$

**Remarks** (2.4.3)

1. Formula (2.4.2) is the covariant generalization of (1.1.8) for the loss of energy-momentum.
2. In the rest frame $\dot{z}(s_1) = (1, 0, 0, 0)$ only energy is lost, and

$$\frac{dE}{dt} = \frac{e^2}{4\pi}\frac{2}{3}\ddot{z}^2.$$

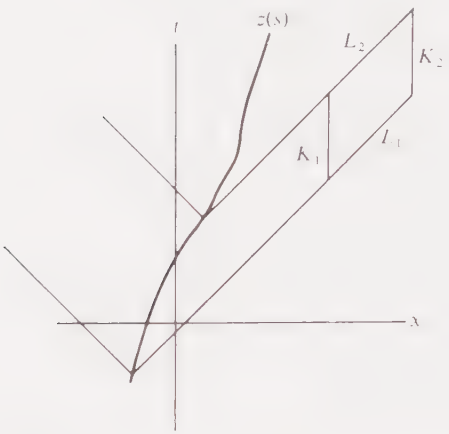



Figure 32 How $N$ is allowed to increase in (2.4.3; 3)

3. The reason $\int_{\partial N \cap K} {}^*\mathcal{F}_z$ is asymptotically independent of $R$ for large $R$ is that $\int_{L_1} {}^*\mathcal{F}_\alpha^{(\vec{z})}$ vanishes, because ${}^*(x - z)_{|L_{1,2}} = 0$ (Problem 1). (See Figure 32.)

   Since the equation $d{}^*\mathcal{F}_\alpha = 0$ is valid away from the world-line $z(s)$,

$$\int_{K_1} {}^*\mathcal{F}_\alpha^{(\vec{z})} - \int_{K_2} {}^*\mathcal{F}_\alpha^{(\vec{z})} = \int_{L_1} {}^*\mathcal{F}_\alpha^{(\vec{z})} - \int_{L_2} {}^*\mathcal{F}_\alpha^{(z)} = 0.$$

**Examples** (2.4.4)

1. Hyperbolic motion. If

   $$z(s) = a^{-1}(\text{Sinh } as, 0, 0, \text{Cosh } as) \quad \text{and} \quad \dot{z}(s) = (\text{Cosh } as, 0, 0, \text{Sinh } as),$$

   then

$$\frac{dP^\alpha}{ds} = \frac{3}{2}\frac{e^2}{4\pi}\dot{z}^\alpha a^2.$$

   The energy contained in the radiation field always increases, while momentum is transferred to it in the negative $z$-direction when $s < 0$, and in the positive $z$-direction when $s > 0$. Therefore the charge always radiates forward along its direction of motion.
2. Synchrotron radiation. With the current of a rotating charge (2.3.14),

$$\dot{z} = \frac{1}{\sqrt{1 - v^2}}(1, -v \sin \omega t, v \cos \omega t, 0),$$

$$\ddot{z} = \frac{-v^2}{R(1 - v^2)}(0, \cos \omega t, \sin \omega t, 0),$$

where $\omega = v/R$. As a consequence, $\delta E$, the energy loss per period $\cdot 2\pi$, is

$$\delta E = \frac{2}{3\omega}\frac{e^2}{4\pi}\ddot{z}^2 = \omega\frac{e^2}{6\pi}\frac{v^2}{(1-v^2)^2}.$$

If we use the value $e^2/4\pi\hbar = \frac{1}{137}$, this is $\sim \omega\hbar(v^2/200)(E/m)^4$. So long as $E/m \sim 1$, the charge clearly needs to undergo more than 200 revolutions to give off a quantum at the ground-state frequency. If $v \to 1$, the rate of energy loss increases rapidly; for, e.g., 5 GeV electrons, $(E/m)^4 \sim 10^{16}$. Accordingly, fast electrons moving in a circle lose quite a bit of energy in the form of synchrotron radiation.

Although $\mathscr{T}_x^{(\xi)}$ does not contribute to the integral over the light-cones $L_1$ and $L_2$ in the calculation of the right side of (2.4.1), the other terms in $\mathscr{T}_x$ have infinite integrals! They decrease as $r^{-3}$ and $r^{-4}$, which diverge when integrated over all space. In order to isolate the causes of the problem, we write

$$d^*\mathscr{T}_\alpha = i_\alpha(F^{\text{in}} + \tfrac{1}{2}F^{\text{rad}} + \tfrac{1}{2}(F^{\text{ret}} + F^{\text{adv}})) \wedge {}^*J \tag{2.4.5}$$

(cf. (2.2.21)). The first term is the Lorentz force from the incoming field, and causes no trouble. We discovered that if the current in (2.2.21) had compact support, then aside from $F^{\text{in}}$ only the radiation field $F^{\text{rad}}$ contributes to the loss of energy-momentum in the limit of infinite times. For this reason, we next evaluate the contribution of $F^{\text{rad}}$ to the left side of (2.4.1) for the point charge. It will turn out that this term is finite, and all the difficulties stem from the last term.

Since $J$ is supported on $z(s)$, we must evaluate $F^{\text{rad}}$ on the world-line of the charge. After some simple algebra (Problem 2), we can write the radiation field as

$$F^{\text{rad}}_{\alpha\beta}(x) = e\int_{-\infty}^{\infty} ds\, D(x - z(s))\frac{d}{ds}\left[\frac{\dot{z}_\alpha(x-z)_\beta}{\langle \dot{z}\,|\,x-z\rangle} - (\alpha \leftrightarrow \beta)\right]. \tag{2.4.6}$$

We wish now to let $x$ carefully approach some point of the world-line, say $z(0)$. To do this, we expand the integrand of (2.4.6) about $s = 0$:

$$z(s) - z(0) = s\dot{z} + \frac{s^2}{2}\ddot{z} + \frac{s^3}{6}\dddot{z} + \ldots, \qquad \dot{z} \equiv \dot{z}(0), \quad \text{etc.}$$

$$\dot{z}(s) = \dot{z} + s\ddot{z} + \frac{s^2}{2}\dddot{z} + \ldots, \tag{2.4.7}$$

and call $x - z(0) = \lambda$. In order that $x$ stay between the retarded time and the advanced time (Figure 33), let $\langle \dot{z}\,|\,\lambda\rangle = 0$ thus $\lambda$ is spacelike as it approaches 0. Since

$$D(x) = \frac{\delta(x^2)}{2\pi}(\Theta(x^0) - \Theta(-x^0))$$

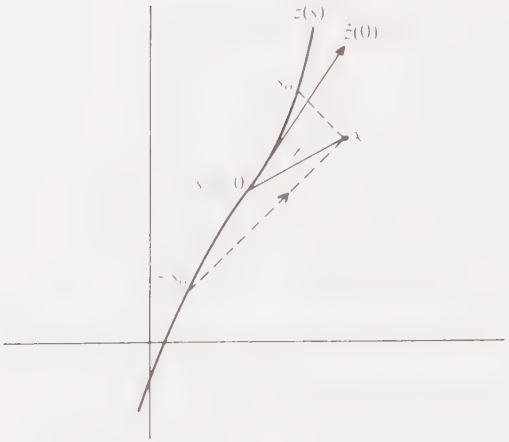


Figure 33 The limit $\lambda \to 0$

and

$$(x - z(s))^2 = \left(\lambda - s\dot{z} - \frac{s^2}{2}\ddot{z}\right)^2 + O(s^3) = \lambda^2 - s^2(1 + \langle\lambda|\ddot{z}\rangle) + O(s^3),$$

if $\lambda \to 0$, then

$$D(x - z(s)) = \frac{\delta(\lambda^2 - s^2)}{2\pi}(\Theta(s) - \Theta(-s)) = \frac{1}{4\pi\lambda}(\delta(s - \lambda) - \delta(s + \lambda)),$$

$$\lambda \equiv \langle\lambda|\lambda\rangle^{1/2} > 0. \tag{2.4.8}$$

If the rest of the integrand of (2.4.6) is also expanded about $s = 0$,

$$\begin{aligned} [\quad] &\equiv \frac{N(\lambda)}{s} + A(\lambda) + sB(\lambda) + \frac{s^2}{2}C(\lambda) \\ \frac{d}{ds}[\quad] &= -\frac{N(\lambda)}{s^2} + B(\lambda) + sC(\lambda), \end{aligned} \tag{2.4.9}$$

then as $\lambda \to 0$, (2.4.6) becomes equal to

$$\frac{e}{4\pi}\lim_{\lambda\to 0}\frac{1}{\lambda}\left(\frac{d}{ds}[\quad]_{|s=\lambda} - \frac{d}{ds}[\quad]_{|s=-\lambda}\right) = \frac{e}{4\pi}2C(0).$$

Substitution of the series (2.4.7) reveals that

$$C(0) = \tfrac{2}{3}(\dddot{z}_\alpha \dot{z}_\beta - \ddot{z}_\beta \ddot{z}_\alpha),$$

with which we obtain a formula for the

**Radiation Field on the World-Line** (2.4.10)

$$\tfrac{1}{2}F^{\text{rad}}(z(s)) = \frac{e}{3\pi}\ddot{z}(s) \wedge \dot{z}(s).$$

**Remarks** (2.4.11)

1. The radiation field is purely electric in the rest-frame of the particle:

$$\tfrac{1}{2}(\mathbf{E}^{\text{ret}} - \mathbf{E}^{\text{adv}}) = \frac{e}{4\pi}\frac{2}{3}\ddot{\mathbf{z}}, \qquad \tfrac{1}{2}(\mathbf{B}^{\text{ret}} - \mathbf{B}^{\text{adv}}) = 0.$$

2. In the result (2.4.10) it is understood that the term $-\ddot{\mathbf{z}}/r$ in $\mathbf{E}$ of (2.3.9) contributes the limit

$$\lim_{r\to 0} \frac{-\ddot{\mathbf{z}}(s-r) + \ddot{\mathbf{z}}(s+r)}{2r} = \dddot{\mathbf{z}}$$

to $F^{\text{rad}}$.

When averaged over space, $(\mathbf{x}(\mathbf{x}\cdot\ddot{\mathbf{z}}))\,r^3$ in like manner contributes $-\ddot{\mathbf{z}}\,3$, and the Coulomb field of the near zone and the magnetic field disappear from (2.3.9). Formula (2.4.10) should a priori be averaged over the different possible directions from which $z(s)$ is approached, but we have noted above that it is independent of the direction of $\lambda$ provided that $\langle \dot{z} | \lambda \rangle = 0$.

If we use (2.4.10) in Equation (2.4.5), and conclude from $\langle \dot{z} | \ddot{z} \rangle = 0$ that $\langle \dot{z} | \dddot{z} \rangle = -\ddot{z}^2$, then we get a formula for the

**Energy and Momentum Lost to the Radiation Field** (2.4.12)

$$\frac{dP_\alpha^{\text{rad}}}{ds} \equiv \frac{-d}{2ds}\int_{N_4} {}^*i_J i_\alpha F^{\text{rad}} = \frac{e}{2}\dot{z}^\beta(s)F_{\beta\alpha}^{\text{rad}}(z(s)) = \frac{e^2}{4\pi}\frac{2}{3}(\dot{z}_\alpha \ddot{z}^2 - \dddot{z}_\alpha).$$

**Remarks** (2.4.13)

1. The limit $ds \to 0$ is understood in (2.4.12), and $N_4$ contains the part of $z(s)$ between $s$ and $s + ds$; thus on the right side the derivative of $z$ is taken at the proper time $s$.
2. The first term on the right is precisely the energy-momentum vector that flows to infinity (2.4.2). It would not be possible for this term to be the whole energy-momentum exchange, because $e\dot{z}^\alpha \dot{z}^\beta F_{\alpha\beta}^{\text{rad}}$ must equal zero and not $-(e^2/6\pi)\ddot{z}^2$. The deficiency is made up by the second term.
3. Although the first term has a definite sign because $\ddot{z}^2 \geq 0$, and represents an irretrievable loss of energy, the second is a total differential and contributes nothing to an integral by $ds$, provided that the initial value of $\ddot{z}$ is returned to at the end. This vector represents retrievable energy-momentum stored in the near zone, and for that reason it is not present at infinity.

4. In the rest-frame $\dot{z} = (1, 0, 0, 0)$, $\ddot{z}$ is $(0, \ddot{\mathbf{z}})$, and $\ddot{z}^2 + (\dot{z}|\dddot{z}) = 0 \Rightarrow \dddot{z} = (\ddot{z}^2, \dddot{\mathbf{z}})$. The two terms in the energy loss cancel out, as the particle has no energy to lose. An accelerated particle that is momentarily at rest borrows the energy that it radiates from the field in the near zone.
5. If $\dot{z} = (1/\sqrt{1 - v^2}, \mathbf{v}/\sqrt{1 - v^2})$, then the $\ddot{z}^2$ term of the reaction on the charge acts like a frictional force retarding $\mathbf{v}$, while the term with $\dddot{\mathbf{z}}$ tends to increase the acceleration of the particle. This leads to all sorts of paradoxical consequences, which we shall return to after having discussed the remaining contribution $\sim F^{\text{ret}} + F^{\text{adv}}$.

**Examples** (2.4.14)

1. By taking more derivatives in Example (2.4.4; 1),

$$\ddot{z}(s) = a(\text{Sinh } as, 0, 0, \text{Cosh } as)$$
$$\dddot{z}(s) = a^2(\text{Cosh } as, 0, 0, \text{Sinh } as),$$

we see that the two terms of (2.4.12) exactly cancel, and $dP^{\text{rad}}\, ds = 0$. A charge in hyperbolic motion radiates on credit: the energy is not supplied by the particle, but comes from the near zone. Of course, the debt must be repaid later, once the acceleration stops. For example, if the charge is accelerated from rest to the velocity $v = \text{Tanh } as_0$,

$$\dot{z}(s) = \Theta(-s)(1, 0, 0, 0) + \Theta(s)\Theta(s_0 - s)(\text{Cosh } as, 0, 0, \text{Sinh } as)$$
$$+ \Theta(s - s_0)\left(\frac{1}{\sqrt{1 - v^2}}, 0, 0, \frac{v}{\sqrt{1 - v^2}}\right);$$

and then

$$\ddot{z}(s) = \Theta(s)\Theta(s_0 - s)a(\text{Sinh } as, 0, 0, \text{Cosh } as),$$
$$\dddot{z}(s) = \Theta(s)\Theta(s_0 - s)a^2(\text{Cosh } as, 0, 0, \text{Sinh } as) + \delta(s)(0, 0, 0, 1)$$
$$-\delta(s - s_0)\left(\frac{v}{\sqrt{1 - v^2}}, 0, 0, \frac{1}{\sqrt{1 - v^2}}\right).$$

Hence

$$\dot{z}\ddot{z}^2 - \dddot{z} = -\delta(s)(0, 0, 0, 1) + \delta(s - s_0)\left(\frac{v}{\sqrt{1 - v^2}}, 0, 0, \frac{1}{\sqrt{1 - v^2}}\right).$$

The force on the particle is the negative of the rate of change of the energy-momentum of the field. At first the particle feels a jolt in the direction of the acceleration from $F^{\text{rad}}$, and later the radiation force operates in the opposite direction, and $F^{\text{in}}$ has to pay the energy bill.

2. For the rotating charge of (2.4.4; 2),

$$\dddot{z} = \frac{v^3}{R^2(1 - v^2)^{3/2}}(0, \sin \omega t, -\cos \omega t, 0).$$

The second term of (2.4.12) causes no additional loss of energy in this case, but it does intensify the braking action of the first term, opposing the velocity of the charge:

$$\dot{z}\ddot{z}^2 - \dddot{z} = \frac{v^3}{R^2(1-v^2)^{5/3}}(v, -\sin\omega t, \cos\omega t, 0).$$

The calculation of the last term of (2.4.5) is simply a matter of replacing the difference appearing in (2.4.8) with a sum:

$$D^{\text{ret}}(x - z(s)) + D^{\text{adv}}(x - z(s)) = \frac{1}{4\pi\lambda}(\delta(s-\lambda) + \delta(s+\lambda)). \quad (2.4.15)$$

Hence the expansion (2.4.9) results in a contribution $-N(\lambda)/\lambda^3 + B(\lambda)/\lambda$, where

$$N(\lambda) = \frac{\dot{z}_\alpha \lambda_\beta - \dot{z}_\beta \lambda_\alpha}{1 + \langle \ddot{z} | \lambda \rangle}.$$

And not only is $N(\lambda)\ \lambda^3$ divergent as $\lambda \to 0$, but it also depends on the direction of $\lambda$: If no direction has been singled out in Minkowski space, then the net result must be $\sim \dot{z}_\alpha \ddot{z}_\beta - \dot{z}_\beta \ddot{z}_\alpha$ and the coefficient goes as $1/\lambda$ for dimensional reasons. Actually the next term $B(0)$ is $(\dot{z}_\alpha \ddot{z}_\beta - \dot{z}_\beta \ddot{z}_\alpha)/2$ independently of the direction of $\lambda$. Normally only the latter term is retained while $N$ is swept under the rug by some averaging procedure; by this hocus-pocus,

$$\tfrac{1}{2}(F^{\text{ret}} + F^{\text{adv}})(z(s)) = \frac{e}{4\pi}\ddot{z} \wedge \dot{z} \lim_{\lambda \to 0} \frac{c}{\lambda}. \quad (2.4.16)$$

Equation (2.4.16) is not very well defined, but at any rate the numerical factor $c$ is positive when calculated in this way. Then attempts are made to argue away the resulting indeterminacy in the radiative reaction of the field along the following lines: Suppose that the three contributions of (2.4.5) are combined and set equal to minus the rate of change of the energy-momentum of a particle of mass $m_0$ during the time $ds$. Then there results

$$m_0 \ddot{z}_\beta = e\dot{z}^\alpha F^{\text{in}}_{\alpha\beta} - \frac{e^2}{4\pi}\frac{2}{3}(\dot{z}_\beta \ddot{z}^2 - \dddot{z}_\beta) - \delta m \ddot{z}_\beta,$$
$$\delta m = \lim_{\lambda \to 0} \frac{ce^2}{4\pi\lambda}. \quad (2.4.17)$$

Next one calls $m = m_0 + \delta m$ ("mass-renormalization") and smugly solves the

**Renormalized Equation of Motion** (2.4.18)

$$m\ddot{z}_\beta = e\dot{z}^\alpha F^{\text{in}}_{\alpha\beta} - \frac{e^2}{4\pi}\frac{2}{3}(\dot{z}_\beta \ddot{z}^2 - \dddot{z}_\beta).$$

**Remarks** (2.4.19)

1. The mass $m$ is clearly what would be measured by the inertia of the particle in an external field. In (2.4.18) there are no apparent infinities.
2. It may seem peculiar that in a theory that is invariant under reversal of the motion, something so obviously not invariant under motion reversal as the radiative reaction force should occur. It arises because of the use of $F^{\text{in}}$ and $D^{\text{ret}}$; it gets its sign from the initial conditions, and would have the opposite sign if $F^{\text{out}}$ and $D^{\text{adv}}$ were used. By using $(D^{\text{ret}} + D^{\text{adv}})\,2$, one can even find stationary solutions of the relativistic two-body problem [23] for which nothing at all is radiated. If the system has a finite energy one finds radiation damping for $t \to \pm\infty$.
3. When discussing (2.2.21) we learned that for currents with compact support the total energy-momentum vector transmitted to the field originates in the Lorentz force with $F^{\text{rad}}$. This result is carried over to point particles in (2.4.16), and it can be seen that $(F^{\text{ret}} + F^{\text{adv}})/2$ simply contributes $\delta m(\dot{z}(\infty) - \dot{z}(-\infty))$, which is the change of the energy-momentum of the self-field attached to the particle.
4. The self-field is eliminated in (2.4.18), and only the particle's coordinates appear. To solve the initial-value problem of the total system, $z$, $\dot{z}$, $F$ and $^*F$ must all be known at some time. One might therefore expect that the elimination of $F$ would render it necessary to take the whole previous history, $z(s)$ for $s < 0$, into account. In fact, the only extra quantity that shows up in the limit of a point particle is $\dddot{z}$, and the solution manifold of the Cauchy problem is only increased by dependence on the three parameters $\ddot{\mathbf{z}}$. But even so, the physically acceptable solution manifold will turn out to be of a lower dimension.
5. The question arises of why we did not circumvent the difficulties connected with point particles by using charges spread over some positive volume. Unfortunately, it is not easy to obtain a theory in this way that has local conservation of energy and momentum (cf. (1.3.25; 2)).
6. The result (2.4.18) can be explained as follows: The field of a point charge has an infinite energy,

$$\frac{1}{2}\int (|\mathbf{E}|^2 + |\mathbf{B}|^2) \to \infty,$$

and since the particle carries this energy along with it, it is subjected to an infinite mass increase $\delta m$ by Einstein's principle that energy is mass. In order that $m$ remain finite, it is necessary to start with an infinite negative "bare" mass $m_0$, obviously a dangerous undertaking. The field $E^{(\ddot{z})} = -e\ddot{z}/r$ causes a reverse acceleration, that is, a braking. This quantity was decomposed above into an infinite term $-\delta m\ddot{z}$ and $+\dddot{z}$. The positive sign of the latter part comes from the use of $D^{\text{ret}}$: The particle feels the field that was produced a short time beforehand, and if the contribution $-\delta m\ddot{z}$ from $(D^{\text{ret}} + D^{\text{adv}})\,2$ is subtracted, the net force is in the direction of the

positive rate of change of the acceleration. This tends to make the particle fly on ahead, if the braking term is compensated for to a large extent by an infinite negative inertia. If a particle is accelerated so quickly that the "acceleration force" $\sim \dddot{z}$ is as large as the braking terms $\dot{z}\ddot{z}^2$ and $\ddot{z}$, then the particle takes off under its own steam.

Unfortunately, mathematics can not be fooled by such simple tricks; the difficulties that were swept under the rug show up later as all sorts of paradoxical consequences of Equation (2.4.18).

**Examples** (2.4.20)

1. The run-away solution:

$$\dot{z} = (\mathrm{Cosh}[\tau_0 a e^{s/\tau_0}],\ \mathrm{Sinh}[\tau_0 a e^{s/\tau_0}], 0, 0),$$

$$\tau_0 = \frac{e^2}{6\pi m}, \qquad a \text{ arbitrary},$$

solves (2.4.18) with $F^{\mathrm{in}} = 0$ (Problem 3). The charge suddenly begins to run away ($\tau_0 \sim 10^{-23}$ seconds for electrons). Since $\ddot{z}^2 = a^2 e^{2s/\tau_0}$, it radiates a tremendous amount of energy: this is consistent with conservation of energy because on the one hand an acceleration takes energy away from a particle with a negative mass $m_0$, and on the other hand there is always energy to be tapped from the infinite reservoir of self-energy and pumped into the far zone.

2. A well aimed shot can bring the flight of the particle to a stop. If $F^{\mathrm{in}}$ is such that

$$eE_1(z) = ma\tau_0 \,\mathrm{Cosh}[\tau_0 a]\delta(z^0 - z^3),$$

then (2.4.18) has the solution

$$\dot{z}(s) = \Theta(-s)(\mathrm{Cosh}[\tau_0 a e^{s/\tau_0}],\ \mathrm{Sinh}[\tau_0 a e^{s/\tau_0}], 0, 0) + \Theta(s)(\mathrm{Cosh}[\tau_0 a],\ \mathrm{Sinh}[\tau_0 a], 0, 0)$$

if $z_3(0) = z_0(0)$, $0 = -\dot{z}^3(0) = \dot{z}^2(0)$ (Problem 3). Such behavior is often felt to be acausal, because the particle starts to accelerate before it is brought to its senses by the pulse from $F^{\mathrm{in}}$ (Figure 34).

**Remark** (2.4.21)

Not all the solutions of (2.4.18) are crazy (see Problem 4). Attempts have been made to separate sense from nonsense by imposing special initial conditions (cf. [24]). It is to be hoped that some day the real solution of the problem of the charge-field interaction will look different, and the equations describing nature will not be so highly unstable that the balancing act can only succeed by having the system correctly prepared ahead of time by a convenient coincidence.

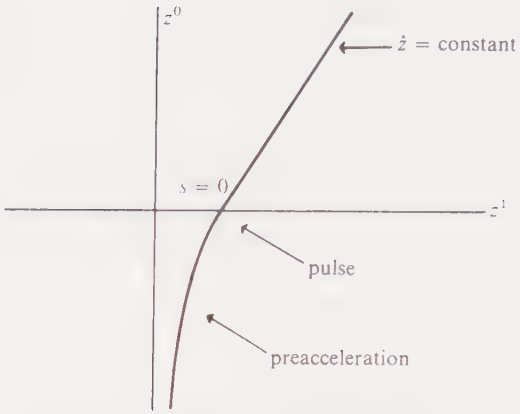



Figure 34 Motion with preacceleration

**Problems** (2.4.22)

1. Verify the claim made in (2.4.3; 3) that $^*(x - z)_{|L_1} = 0$.
2. Derive (2.4.6).
3. Show that the $\dot{z}$'s of Examples (2.4.20; 1) and (2.4.20; 2) solve (2.4.18).
4. Solve (2.4.18) for $F^{\text{in}}$ a constant electric field in the $x_1$-direction and no magnetic field. Use the ansatz $\dot{z}(s) = (\text{Cosh } \omega(s), \text{Sinh } \omega(s), 0, 0)$. Compare with (I: 5.2.19; 3) and (2.4.20; 1).

**Solutions** (2.4.23)

1. Let $z(s) = 0$: $r_{|L_1} = t_{|L_1} \Rightarrow dr_{|L_1} = dt_{|L_1}$. Hence

$$^*x = \mathbf{x} \cdot {}^*d\mathbf{x} - t^*dt = r^*dr - t^*dt = r\, d\Omega \wedge dt - t\, d\Omega \wedge dr\colon {}^*x_{|L_1} = 0.$$

2. Let $y \equiv (\bar{x} - z(s))^2$. Then

$$A^{\text{rad}}_{\alpha,\beta}(\bar{x}) = -e \int ds\, \dot{z}_\alpha \frac{\partial y}{\partial \bar{x}^\beta} \frac{ds}{dy} \frac{d}{ds} D(y) = -e \int ds\, D(\bar{x} - z(s)) \frac{d}{ds} \frac{\dot{z}_\alpha(s)(\bar{x} - z(s))_\beta}{\langle \dot{z}(s) | \bar{x} - z(s) \rangle}.$$

3. We shall verify Example (2.4.21; 2), and the solution of (2.4.20; 1) will appear as a by-product.

$$\ddot{z} = \Theta(-s) a e^{s/\tau_0} (\text{Sinh}[\ ], \text{Cosh}[\ ], 0, 0)$$

$$\dddot{z} = \Theta(-s) a^2 e^{2s/\tau_0} (\text{Cosh}[\ ], \text{Sinh}[\ ], 0, 0) + \ddot{z}/\tau_0 - \delta(s) a (\text{Sinh } \tau_0 a, \text{Cosh } \tau_0 a, 0, 0)$$

$$\dot{z}\ddot{z}^2 - \dddot{z} = -\ddot{z}/\tau_0 + \delta(s) a (\text{Sinh } \tau_0 a, \text{Cosh } \tau_0 a, 0, 0).$$

Now,

$$\delta(z^0(s) - z^3(s)) = \frac{1}{\dot{z}^0(0)} \delta(s) = \frac{\delta(s)}{\text{Cosh } \tau_0 a},$$

and hence $e\dot{z}^\alpha F^{\text{in}}_{\alpha\beta} = \delta(s)\, m\, a\, \tau_0 (-\text{Sinh } \tau_0\, a, \text{Cosh } \tau_0\, a, 0, 0)$ and $m\ddot{z}_\beta = e\dot{z}^\alpha F^{\text{in}}_{\alpha\beta} - m\tau_0(\dot{z}_\beta \ddot{z}^2 - \dddot{z}_\beta)$.

$$\ddot{z} = \dot{\omega}(\text{Sinh } \omega, \text{Cosh } \omega, 0, 0), \qquad \ddot{z}^2 = \dot{\omega}^2,$$
$$\dddot{z} = \dot{\omega}^2(\text{Cosh } \omega, \text{Sinh } \omega, 0, 0) + \ddot{\omega}(\text{Sinh } \omega, \text{Cosh } \omega, 0, 0),$$
$$-\dot{z}\ddot{z}^2 + \dddot{z} = \ddot{\omega}(\text{Sinh } \omega, \text{Cosh } \omega, 0, 0).$$

Equation (2.4.18) requires that $\dot{\omega} = E/m + \tau_0 \ddot{\omega}$, which implies that

$$\omega(s) = a + \frac{E}{m} s + c\tau_0 e^{s/\tau_0},$$

where $a$ and $c$ are constants of integration. Only if $c = 0$, that is, for the special initial condition $\ddot{z}(0)^2 = E^2/m^2$, is there no self-acceleration.

# 3 The Field in the Presence of Conductors

## 3.1 The Superconductor

*The superconductor is a simple model of a coupled system of equations for charged matter and an electromagnetic field. As a perfect conductor and diamagnet it excludes all electric and magnetic fields from its interior.*

Realistic situations do not very closely resemble the idealization discussed in the preceding chapter, where the charge distribution is prescribed. The field in turn influences the motion of the charges, so it would be more correct to analyze the coupled system. For a point-particle the analysis is subject to the difficulties encountered in §2.4. Moreover, the charge-carriers in matter, electrons and atomic nuclei, are governed by the laws of quantum mechanics, and their motion is a very complicated many-body problem. Every phenomenological description of matter is of necessity either highly idealized or else so general as to contain little information. Notwithstanding that objection, in order to formulate the ideas of this chapter mathematically, we shall single out one of the many models for a superconductor, which can be cast in a simple mathematical form. It is good enough for our purposes, as we shall always consider an extreme case in the examples, for which the charge-carriers in matter are numerous and move about freely. By responding instantaneously to any applied field, they cause the net field within the material to disappear entirely. Later, when we treat the gravitational interaction, this model will serve as our prototype of charged matter.

**London's Equations** (3.1.1)

Consider the hydrodynamic equations of an incompressible, charged, frictionless fluid in an electromagnetic field, calling the velocity field $\mathbf{v}(\mathbf{x}, t)$:

$$\frac{d\mathbf{v}}{dt} \equiv \frac{\partial \mathbf{v}}{\partial t} + \nabla \frac{v^2}{2} - [\mathbf{v} \times \nabla \times \mathbf{v}] = \frac{e}{m}(\mathbf{E} + [\mathbf{v} \times \mathbf{B}]). \tag{3.1.2}$$

Then from the equation $\nabla \times \mathbf{E} = -\dot{\mathbf{B}}$ results the generalization of Helmholtz's circulation theorem,

$$\dot{\mathbf{w}} = \nabla \times [\mathbf{v} \times \mathbf{w}], \qquad \text{where} \quad \mathbf{w} \equiv \nabla \times \mathbf{v} + \frac{e}{m}\mathbf{B}. \tag{3.1.3}$$

Therefore, if $\mathbf{w}$ is zero at any time, it is always zero. This means that the curl of $\mathbf{v}$ arises only from the vortices created when $\mathbf{B}$ is switched on (cf. I: §5.4), and the equations simplify to

$$\begin{aligned} \frac{\partial \mathbf{v}}{\partial t} + \nabla \frac{v^2}{2} &= \frac{e}{m}\mathbf{E} \\ \nabla \times \mathbf{v} &= -\frac{e}{m}\mathbf{B}. \end{aligned} \tag{3.1.4}$$

If one now writes

$$\mathbf{J} = \frac{e\rho \mathbf{v}}{\sqrt{1 - v^2}}, \qquad J^0 = \frac{e\rho}{\sqrt{1 - v^2}}, \tag{3.1.5}$$

and if $\rho$ is constant, then

$$J_{\beta,\alpha} - J_{\alpha,\beta} = \frac{\rho e^2}{m} F_{\alpha\beta} \tag{3.1.6}$$

to an accuracy of order $v^2 \ll 1$. This equation together with Maxwell's equations will be the foundation of our model. It admits a coordinate-independent formulation,

$$F = \frac{dJm}{\rho e^2}, \qquad \delta F = J + j. \tag{3.1.7}$$

**Remarks** (3.1.8)

1. The current $j$ consists of charges not participating in the superconductive current $J$. We shall take $j$ as given, and assume that $\delta j = 0$; then (3.1.7) implies that $dF = \delta J = 0$.
2. We shall ignore the heuristic derivation to the point of not requiring that $\langle J | J \rangle = -e^2\rho^2$, which follows from (3.1.5).

3. For the present, $\rho$ is regarded as a constant, known as the density of the superconducting electrons. There will later be a discussion of the variable $\rho$.
4. Equation (3.1.7) shows that the manifold is not provided with any additional structure; for instance, there is no distinguished rest-frame.

**The Integral Form of the Equations** (3.1.9)

If $F = dA$, then (3.1.6) is equivalent to

$$\int_{\partial N_2} \left( J - \frac{\rho e^2}{m} A \right) = 0, \qquad \dim N_2 = 2.$$

Two important special cases are

(i) $N = \{t = z = 0, x^2 + y^2 \leq R^2\}$:

$$\int_{x^2+y^2=R^2} d\mathbf{s} \cdot \mathbf{J} = \frac{\rho e^2}{m} \int_{x^2+y^2=R^2} ds_j A_j = -\frac{\rho e^2}{m} \int_{x^2+y^2<R^2} d\mathbf{S} \cdot \mathbf{B},$$

i.e., the current circulates in proportion to the magnetic flux (cf. Remark (I: 5.1.10; 1), with $A_j$ equal to minus the vector potential).

(ii) $N = \{y = z = 0, t_1 \leq t \leq t_2\}$:

$$\int_{t=t_2} dx\, J_x - \int_{t=t_1} dx\, J_x = \frac{\rho e^2}{m} \int_{t_1}^{t_2} dt\, dx\, E_x.$$

The rate of change of the superconductive current is given by the integral over the electric field.

**The Elimination of the Superconductive Current** (3.1.10)

From (3.1.7) we get a second-order equation for $F$,

$$\left( -\Delta + \frac{\rho e^2}{m} \right) F = -dj. \tag{3.1.11}$$

The solution of (3.1.11) requres a Green function satisfying

$$\left( -\Delta + \frac{\rho e^2}{m} \right) G^{\text{ret}}_{\bar{x}} = \delta_{\bar{x}}, \tag{3.1.12}$$

with which

$$\begin{aligned} F(\bar{x}) &= F^{\text{in}}(\bar{x}) - \int G^{\text{ret}}_{\bar{x}} \wedge dj, \\ \left( -\Delta + \frac{\rho e^2}{m} \right) F^{\text{in}} &= 0. \end{aligned} \tag{3.1.13}$$

We have assumed that $j$ decreases sufficiently fast at infinity, so that the integrals can extend over the whole manifold as in (2.2.15), without any boundary terms.

In Minkowski space $(\mathbb{R}^4, \eta)$ it is easy to construct a Green function satisfying (3.1.12). As in (2.2.3) it has the form

$$G_{\bar{x}}^{\text{ret}} = \tfrac{1}{2}\bar{e}_{\alpha\beta} \otimes {}^*e^{\alpha\beta}\Delta^{\text{ret}}(\bar{x} - x), \tag{3.1.14}$$

where

$$\Delta^{\text{ret}}(x) = (2\pi)^{-4} \int d^4k \, e^{i\langle k|x\rangle}\left(k^2 + \frac{e^2\rho}{m}\right)^{-1}. \tag{3.1.15}$$

The integration path for $k^0$ again passes above the poles at $\pm\sqrt{|\mathbf{k}|^2 + e^2\rho/m}$, as in Figure 18, in order that $\Delta^{\text{ret}}(x) = 0$ for $x^0 < |\mathbf{x}|$. The integral (3.1.15) can be expressed in terms of Hankel functions [22], and if $\Delta^{\text{ret}}$ is integrated over time, the result is a Yukawa potential,

$$\int_{-\infty}^{\infty} dt \, \Delta^{\text{ret}}(x) = \frac{e^{-r(\rho e^2/m)^{1/2}}}{4\pi r} \tag{3.1.16}$$

(Problem 1). If (3.1.13) is written as

$$\begin{aligned} F &= F^{\text{in}} + dA^{\text{ret}} \\ A_\alpha^{\text{ret}}(\bar{x}) &= -\int d^4x \, \Delta^{\text{ret}}(\bar{x} - x) j_\alpha(x), \end{aligned} \tag{3.1.17}$$

then in the static limit,

$$A_\alpha^{\text{ret}}(\bar{\mathbf{x}}) = -\int \frac{d^3x}{4\pi|\mathbf{x} - \bar{\mathbf{x}}|} \exp\left(-|\mathbf{x} - \mathbf{x}|\left(\frac{\rho e^2}{m}\right)^{1/2}\right) j_\alpha(x). \tag{3.1.18}$$

**Remarks** (3.1.19)

1. A bounded solution $\sim \exp(i\langle k|x\rangle)$ for $F^{\text{in}}$ exists only for $(k^0)^2 = |\mathbf{k}|^2 + \rho e^2/m \geq \rho e^2/m$. The significance of the **plasma frequency** $\sqrt{\rho e^2/m}$ is evident in the following electrostatic situation: Suppose charges $e$ are arrayed along a line at the points $nL$, $n \in \mathbb{Z}$. If one charge is displaced slightly from equilibrium by $x \ll L$ (see Figure 35), then it feels a force $e^2((L + x)^{-2} - (L - x)^{-2}) \sim -e^2 x L^{-3}$ from its two nearest neighbors. If this is set equal to $m\ddot{x}$, the equation is oscillatory at the frequency $(e^2\rho/m)^{1/2}$, where we have identified the density $\rho$ as $L^{-3}$. The oscillations are associated with solutions having $\mathbf{k} = \mathbf{0}$ and $k^0 = (e^2\rho/m)^{1/2}$.

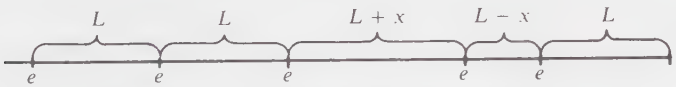



Figure 35 A chain of oscillating charges

2. There are no nontrivial static solutions ($k^0 = 0$) for $F^{\text{in}}$. According to Equation (3.1.18), the field of a static charge does not penetrate a superconductor, but decreases exponentially within a skin-depth also given by $(e^2\rho/m)^{1/2}$. The cause is the induced current $J$, which can be calculated from (3.1.7) with the boundary condition $F^{\text{in}} = 0$ and (3.1.18) as

$$J_\alpha(\bar{\mathbf{x}}) = -\int \frac{d^3x\, e^{-|\mathbf{x}-\bar{\mathbf{x}}|(\rho e^2/m)^{1/2}}}{4\pi|\mathbf{x}-\bar{\mathbf{x}}|} j_\alpha(\mathbf{x})\frac{\rho e^2}{m}$$

in the static case; it opposes the original current and completely cancels $j$ at large distances:

$$\int d^3x\, j_\alpha(\mathbf{x}) = -\int d^3x\, J_\alpha(\mathbf{x}).$$

3. The connection between the Fourier transforms of the external and induced currents with the appropriate definition of $(\ldots)^{-1}$ is

$$\tilde{J} = -\left(1 + \frac{mk^2}{\rho e^2}\right)^{-1}\tilde{j}.$$

If $k = 0$, they are equal and opposite. The 2-form $G$ with Fourier transform

$$\tilde{G} = \left(1 + \frac{\rho e^2}{mk^2}\right)\tilde{F}$$

satisfies Maxwell's equations with no superconducting current,

$$dG = 0, \qquad \delta G = j.$$

Since $\rho$ is constant, the fields $\mathbf{D}$ and $\mathbf{H}$ of phenomenological electrodynamics satisfy these same equations, where the factor $1 + \rho e^2/mk^2$ corresponds on the one hand to a dielectric constant

$$\varepsilon(k) = 1 + \frac{\rho e^2}{mk^2}, \qquad \tilde{\mathbf{D}} = \varepsilon\tilde{\mathbf{E}};$$

and on the other to a magnetic susceptibility

$$\kappa(k) = \frac{-1}{1 + \dfrac{k^2 m}{\rho e^2}}, \qquad \tilde{\mathbf{B}} = (1 + \kappa)\tilde{\mathbf{H}}.$$

Both $\varepsilon(k)$ and $\kappa(k)$ are Lorentz-invariant and commute as convolution operators with translations, but they depend on the frequency, because the relationship between $F$ and $G$ is not local. If $k = 0$, then $\varepsilon = \infty$ and $\kappa = -1$; $\mathbf{E}$ and $\mathbf{B}$ become zero within the material, as they are shielded in a perfect conductor and diamagnet. The theory with the field $J$ does not distinguish a rest frame and the velocity of light is unchanged. Thus $c = 1 = \varepsilon(1 + \kappa)$ and $\kappa \to -1 \quad \varepsilon \to \infty$. Thus in a Lorentz invariant situation charge screening and expulsion of the magnetic field go together.

4. In our case $G$ is just the field generated by the current $j$. In more realistic modes the connection between $F$ and $G$ is tensorial and nonlocal, even when nonlinear effects are neglected. If

$$G_{ik}(x) = \int d^4x' K_{iklm}(x, x')F^{lm}(x'),$$

then $dG \neq 0$ and $G$ no longer satisfies Maxwell's equations with the external current.

Now that we have a useful system of equations for the field plus the charges, let us take the opportunity to study the conservation theorems for the total system, by using the

**Lagrangian Formulation** (3.1.20)

*If we particularize* (3.1.7) *by setting*

$$F = dA, \qquad J = \frac{e\rho}{m}(dS + eA), \qquad S \in E^0,$$

*then the Lagrangian*

$$\mathscr{L} = -\frac{1}{2}\frac{m}{\rho e^2} J \wedge {}^*J - \frac{1}{2} F \wedge {}^*F$$

*reproduces Equations* (3.1.7) *without j.*

**Proof**

Making a variation of $\mathscr{L}$ as in (2.1.2) yields

$$\delta\mathscr{L} = \frac{\delta S\, d^*J}{e} - d\left(\frac{\delta S^*J}{e}\right) - \delta A \wedge [d^*F + {}^*J] - d[\delta A \wedge {}^*F]$$

(where $\delta$ is the variation, not the codifferential). Hence the Euler-Lagrange equations are

$$d^*J = 0, \qquad d^*F = -{}^*J. \qquad \square$$

**Remarks** (3.1.21)

1. Even the "superpotential" $S$ must be changed if a gauge transformation is made; if $F$ and $J$, and consequently $\mathscr{L}$, are to be invariant under $A \to A + d\Lambda$, $S \to S - e\Lambda$.
2. It is possible to express $J \wedge {}^*J$ as $\overline{(d + ieA)\varphi} \wedge {}^*(d + ieA)\varphi$ by use of a complex field $\varphi = \exp(iS)$. The effect of $A$ in this scheme is to make the exterior differential invariant under $\varphi \to \exp(ie\Lambda(x))\varphi$. (See §4.1.)

3. The scalar model discussed so far turned out to be of more importance in elementary particle physics than in condensed matter physics. It showed that in a gauge invariant theory a "mass term" can appear in Maxwell's equations. Exponentially decaying Green functions correspond in quantum theory to particles with a mass and one stumbling block in the unification of weak interactions with electromagnetism was that the particles mediating the weak interaction did not seem to be massless. In fact, they were now found to be exceedingly heavy which is believed to come from their interaction with a scalar field.

**The Energy-Momentum Forms** (3.1.22)

According to Noether's theorem (2.1.5), the 3-forms

$$\frac{(L_v S)^* J}{c} + (L_v A) \wedge {}^*F + i_v \mathscr{L} = \frac{m}{2\rho c^2}(J \wedge i_v {}^*J + (i_v J)^* J) + \tfrac{1}{2}((i_v F) \wedge {}^*F - F \wedge i_v {}^*F) + d((i_v A)^* F)$$

are closed for all Killing vector fields $v$ (cf. Definition (2.1.9)).

**Remarks** (3.1.23)

1. The gauge-dependent term, $d((i_v A)^* F)$, has again shown up. Since it is exact, the rest of the right side of (3.1.22) must be closed. As will be seen in (4.2.9), only that part interacts with the gravitational field, and will be used as the energy-momentum tensor.
2. All the generators of the Poincaré group can function as $v$. However, since $J$ is a 1-form, the presence of matter breaks the conformal invariance (see (2.1.10; 4)); the skin-depth is a distinguished length.
3. If $v$ is the generator $\partial_\alpha$ of a translation, then the 3-form (3.1.22) gets a contribution $t_{\alpha\beta}\, dx^\beta$ from the matter, where

$$t_{\alpha\beta} = \frac{m}{\rho c^2}[J_\alpha J_\beta - \tfrac{1}{2}\eta_{\alpha\beta} J_\gamma J^\gamma].$$

**The Properties of the Energy-Momentum Tensor of Matter** (3.1.24)

(a) $t_{\alpha\beta} = t_{\beta\alpha}$;

(b) $t_{00} = \dfrac{m}{2\rho e^2}[J_0^2 + |\mathbf{J}|^2] \geq 0$, and $= 0$ only if $J = 0$;

(c) $t_{0i} = \dfrac{m}{\rho e^2} J_0 J_i, \qquad {t_{00}}^2 \geq \sum_{i=1}^{3} {t_{0i}}^2.$

**Remarks** (3.1.25)

1. Property (a) follows from Lorentz invariance; but $t_\alpha{}^\alpha \neq 0$, because of the lack of conformal invariance (cf. (2.1.11)).
2. Since there is the same positivity property as in (2.1.13) for the electromagnetic energy-momentum tensor, the argument of (2.1.16) about the uniqueness of the Cauchy problem is still applicable.
3. A relativistic fluid is described phenomenologically by the field $u \in E_1$ of its four-velocity ($\langle u|u\rangle = -1$), its mass-density $\bar{\rho}$, and its pressure $p$. The energy-momentum tensor $T_{\alpha\beta} = (\bar{\rho} + p)u_\alpha u_\beta + p g_{\alpha\beta}$ is by construction diagonal in the rest system ($u = (1, 0, 0, 0)$) and in no other. Its eigenvalues are $(\rho, p, p, p)$. By a comparison with (3.1.23; 3) one arrives at the identification

$$u_\alpha = J_\alpha (J_0^2 - |\mathbf{J}|^2)^{-1/2}, \qquad \bar{\rho} = p = \frac{1}{2}\frac{m}{\rho e^2}(J_0^2 - |\mathbf{J}|^2).$$

   In normal matter, $p \sim$ the density of kinetic energy $\sim 10^{-10}\bar{\rho}$, while for radiation $p = \bar{\rho}/3$. Thus the pressure is unrealistically high in this model.
4. If $e = 0$, then $S$ satisfies d'Alembert's equation $\Delta S = 0$. The compressional waves, which may be thought of as sound, therefore propagate at the speed of light, as is to be expected from $\partial p/\partial\bar{\rho} = 1$. Thus $S$ describes a fluid that is as incompressible as possible without allowing sound to travel faster than light.
5. The reader may be wondering what happens to these sound waves if $e \neq 0$. In that case $S$ loses its physical significance, and can be made to disappear by a gauge transformation. It turns out that the sound waves then reappear as a longitudinal oscillatory mode of $A$, and if $e = 0$, then $A$ can oscillate only transversally (Problem 3).

In practically important problems, one part of space usually contains the metal and all the rest is a vacuum. In that case $\rho$ is not a constant in the model, but instead it changes discontinuously to zero at the metallic surface. A slowly varying field with frequency $k^0 \ll (\rho e^2/m)^{1/2}$ would consequently not penetrate the metal, but would decrease exponentially at the surface within a depth $(\rho e^2/m - (k^0)^2)^{-1/2}$. For simplicity we consider the limit $\rho e^2/m \to \infty$, thus disengaging ourselves from the details of the model. The only essential feature that is preserved is that the field $F$ is excluded from within the metal. It is screened by a surface current $J$, which has a delta-function singularity at the surface, in the limit $\rho e^2/m \to \infty$ (cf. (3.1.19; 2)). The equations (3.1.7) of the model are replaced with the

**Metallic Boundary Conditions** (3.1.26)

*Suppose that the four-dimensional submanifold $CN$ is filled with metal, so that $F = 0$ on it, and let the surface $\partial N$ be given locally by the equation $u = 0$.*

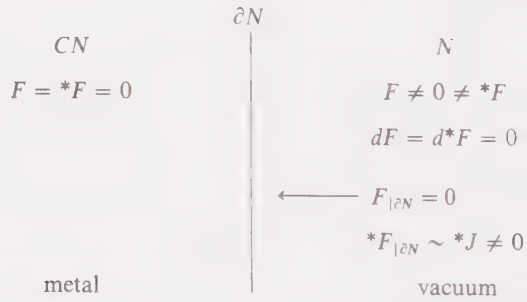
*Then the restriction* (1.2.7: 3) $F_{|\partial N}$ *of the field to* $\partial N$ *must vanish, and* $F$ *is screened by the surface current* $\delta(u)(i_{du}F)_{|\partial N}$.

**Proof**

It is only necessary to extend the arguments of (2.2.1) to allow the current to have a delta-function singularity. If $F = \Theta(u)F'$, where $F'$ is continuous, then $0 = dF = \delta(u)du \wedge F'$ and $-{*}J = d{*}F = \delta(u)du \wedge {*}F'$, where we consider only the singular parts. Because of the $\delta(u)$, it only matters what $F'$ is at $u = 0$, and all the terms containing a factor $du$ drop out of the exterior product. What is left over is simply the restriction of $F'$ according to (1.2.7: 3), and so we conclude from the first of the two equations above that $F'_{|\partial N} = 0$. If we make use of $du \wedge {*}F = -{*}i_{du}F$ in the second equation, the claim made about the surface current follows. □

**Remarks** (3.1.27)

1. If $u = x_1$, making $N = \{x_i \in \mathbb{R}: x_1 > 0\}$, then $dx_{1|\partial N} = 0$ and $F_{|\partial N} = (E_2\, dt \wedge dx_2 + E_3\, dt \wedge dx_3 - B_1\, dx_2 \wedge dx_3)_{|\partial N}$. Therefore $E_2$, $E_3$, and $B_1$ must vanish. The interpretation of this is that surface charges do not produce any discontinuous tangential components of the electric field, and surface currents do not produce discontinuous normal components of the magnetic field.
2. The situation is drawn schematically below:




Since the surface current is not specified beforehand, it may be asked how the initial-value problem is to be solved. While the general solution (1.2.36) is always valid, it is not immediately useful, since $*F$ occurs in the surface integral as well as $F$. It would seem to be necessary to know the restrictions of both to the surface, and we only know that $F_{|\partial N} = 0$. If we manage to find a $G_{\vec{x}}$ such that $*dG_{\vec{x}|\partial N} = 0$, however, then there are no unknown surface contributions, and the solution works as in Chapter 2. In other words, the key to the problem is

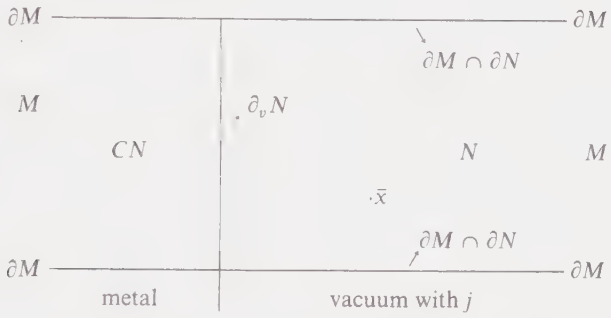
**The Green Function for Metallic Boundary Conditions** (3.1.28)

Let $M \subset \mathbb{R}^4$ be a part of Minkowski space bounded by spacelike hypersurfaces $\partial M$. Suppose that metal fills $CN \cap M$ and that the current $j$ is known in $N \subset M$ and $\partial_v N \equiv \partial N \setminus \partial N \cap \partial M$ is the vertical boundary of $N$. If $G_{\bar{x}}$ satisfies the equations $-\Delta G_{\bar{x}} = \delta_{\bar{x}}$ and $*dG_{\bar{x}|\partial_v N} = 0 \; \forall \bar{x} \in N \setminus \partial N$, then for all $x \in N \setminus \partial N$ the field strength is given by

$$F(\bar{x}) = \int_N dG_x \wedge j - \int_{\partial M \cap N} [\delta G_x \wedge F - *dG_{\bar{x}} \wedge *F].$$

**Remarks** (3.1.29)

1. The situation looks as follows:




   In (3.1.28) it is only relevant to know $F$ and $*F$ on the initial and final surfaces $\partial M \cap \partial N$, and not on $\partial_v N$. In later examples $G_x$, like $G_x^{\text{ret}}$, will be zero outside the past light-cone of $\bar{x}$, and thus only the initial surface affects the integral. The reason this $G_{\bar{x}}$ is selected is that it expresses $F$ in terms of the initial data, and it automatically takes care of the effect of the currents in the upper surface.
2. Strictly speaking, $N$ is not a manifold with a boundary, because it has a sharp edge. But since the integration by parts used in (1.2.36) can also be justified on regions of the form $\{(x, y) \in \mathbb{R}^2 : x \geq 0, y \geq 0\}$, this presents no real obstacle.
3. The $G_{\bar{x}}$ of (3.1.28) is not uniquely determined. However, as long as we possess some $G_x$ that vanishes outside the past light-cone, the formula (3.1.28) guarantees the uniqueness of the Cauchy problem.

In the following sections we shall prove the existence of the $G_{\bar{x}}$ used in (3.1.28) by explicit construction, making use of the well-known method of images from electrostatics. This static method is generalized by means of $G_{\bar{x}}$ for charges in arbitrary motion.

**Problems** (3.1.30)

1. Calculate the integral (3.1.16).
2. Show that $t_{\alpha\beta,}{}^{\beta} = J_\beta F^\beta{}_\alpha$, with $t_{\alpha\beta}$ as in (3.1.23; 3).
3. Show that the equations

$$dJ = \frac{e^2\rho}{m} F, \qquad \delta F = J, \qquad \frac{e^2\rho}{m} > 0,$$

have three linearly independent solutions $\sim \exp(i\langle k|x\rangle)$ if

$$k^2 = -\frac{e^2\rho}{m},$$

and otherwise have only the trivial solution $F = 0$.

**Solutions** (3.1.31)

1. If $\mu = \sqrt{e^2\rho/m}$, then

$$\int_{-\infty}^{\infty} dt(2\pi)^{-4} \int d^4k\, e^{i\langle k|x\rangle}(k^2 + \mu^2)^{-1} = (2\pi)^{-3} \int d^3k\, e^{i\mathbf{k}\cdot\mathbf{x}}(|\mathbf{k}|^2 + \mu^2)^{-1}$$

$$= (2\pi)^{-2} \int_0^{\infty} \frac{k^2\,dk}{k^2 + \mu^2} \int_{-1}^{1} d\eta\, e^{ik r\eta}$$

$$= (2\pi)^{-2} \int_{-\infty}^{\infty} \frac{k\,dk}{k^2 + \mu^2} \frac{e^{ikr}}{ir} = \frac{e^{-\mu r}}{4\pi r}.$$

2. Because of (3.1.6) and $J_{\beta,}{}^{\beta} = 0$,

$$t_{\alpha\beta}{}^{,\beta} = \frac{m}{\rho e^2}\left[J_\alpha{}^{,\beta} J_\beta - J^\beta{}_{,\alpha} J_\beta\right] = F_{\beta\alpha} J^\beta.$$

3. In the Fourier-transformed space the equations become

$$k_\mu k^\rho \tilde{F}_{\rho\nu} - k_\nu k^\rho \tilde{F}_{\rho\mu} = \frac{e^2\rho}{m} \tilde{F}_{\nu\mu}.$$

Multiplying this by $\cdot k^\nu$,

$$-k^2 k^\rho \tilde{F}_{\rho\mu} = \frac{e^2\rho}{m} k^\nu \tilde{F}_{\nu\mu}.$$

Since $\tilde{A}_\mu - k_\mu \tilde{S} \equiv -(m/e^2\rho)k^\rho \tilde{F}_{\rho\mu}$ vanishes iff $F \equiv 0$, either $\tilde{F}_{\nu\mu} = 0$ or else $k^2 = -e^2\rho/m$. Then $\tilde{F}$ is of the form $\tilde{F}_{\nu\mu} = k_\nu \tilde{A}_\mu - k_\mu \tilde{A}_\nu$, which vanishes only if $\tilde{A}_\nu \sim k_\nu$. Thus for the three directions other than $k$ there are nonvanishing solutions.

## 3.2 The Half-Space, the Wave-Guide, and the Resonant Cavity

*The general solution of Maxwell's equations with metallic boundary conditions is easy to construct for simple geometric arrangements of the conductors.*

Classically, the electromagnetic problem in the presence of metallic surfaces is usually conceived as the quest for particular solutions. Here we shall proceed directly to the more general problem, and solve the Cauchy problem by specifying the $G_{\bar{x}}$ of (3.1.28). The interesting question will be what the causal structure of the Green function is. Like the $G^{\text{ret}}_{\bar{x}}$ of (2.2.7), its support will be restricted to the full past light-cone of $\bar{x}$, but unlike $G^{\text{ret}}_{\bar{x}}$ not to its surface. This fact is due to waves that reflect from the metallic surfaces and return at some later time. Such echoes may apparently violate causality, as when they give rise to phase velocities greater than the speed of light. In all the problems discussed below we replace the conducting material with a metallic boundary condition. The currents induced on the conductors do not appear, so we may use the symbol $J$ for the externally prescribed current.

We begin with a trivial warm-up exercise, the problem of a plane metallic mirror. The method of solution introduced points the way to the procedure for more complicated problems.

**The Half-Space** (3.2.1)

*In the notation of* (3.1.28), *let*

$$M = \{x^i \in \mathbb{R}^4 : x^0 \geq t^0\},$$
$$N = \{x^i \in M : x^1 \geq 0\}.$$

*The symbol R will stand for the reflection* $(x^0, x^1, x^2, x^3) \to (x^0, -x^1, x^2, x^3)$ *in M and at the same for the induced mapping on the space of tensors* (I: 2.4.19). *Then*

$$G_{\bar{x}} = (1 + R)G^{\text{ret}}_{\bar{x}},$$

*with* $G^{\text{ret}}_{\bar{x}}$ *from* (2.2.7), *is the Green function needed for this problem.*

**Proof**

The rules for manipulating the diffeomorphism $R$ of the space of vectors are such that it can be interchanged with sums, products, and exterior differentiation:

$$R(\omega + \nu) = R\omega + R\nu, \qquad R(\omega \wedge \nu) = R\omega \wedge R\nu, \qquad \omega, \nu \in E,$$
$$R(e^0, e^1, e^2, e^3) = (e^0, -e^1, e^2, e^3), \qquad e^i = dx^i.$$

However, $R$ reverses the orientation so that $R^*\omega = -^*R\omega$, $\forall\omega \in E_p$ (cf. (1.2.17; 1)).

We now make use of a

**Lemma**

*If $\omega \in E_p$ and $\omega = -R\omega$, then $\omega_{|\partial_v N} = 0$.*

**Proof of Lemma**

Let

$$\omega = \sum_{(i)} \omega_{i_1 \cdots i_p}(x) e^{i_1 \cdots i_p}, \qquad e^i = dx^i, \qquad \omega_{i_1 \cdots i_p} \in E_0.$$

If $i_1 \cdots i_p$ contains the index 1, then $e^{i_1 \cdots i_p}{}_{|x^1=0} = 0$, because $dx_{1\,|x_1=0} = 0$. If $i_1 \cdots i_p$ does not contain the index 1, then $\omega_{i_1 \cdots i_p}(x^0, -x^1, x^2, x^3) = -\omega_{i_1 \cdots i_p}(x^0, x^1, x^2, x^3)$, which must vanish when $x^1 = 0$.

This lemma implies the property $^*dG_{\bar{x}\,|\partial_v N} = 0$ required in (3.1.28), because

$$R^*d(1 + R)G^{\text{ret}}_{\bar{x}} = -^*dR(1 + R)G^{\text{ret}}_{\bar{x}} = -^*d(1 + R)G^{\text{ret}}_{\bar{x}}.$$

Observe that none of the operations of this equation affect $\bar{x}$, and that $R^2 = 1$. The property $-\Delta G_{\bar{x}} = \delta_{\bar{x}}$ $\forall\bar{x} \in N$ is proved by noting that the factor $\delta((\bar{x} - x)^2)$ has been replaced with $\delta((\bar{x}^0 - x^0)^2 - (\bar{x}^1 + x^1)^2 - (\bar{x}^2 - x^2)^2 - (\bar{x}^3 - x^3)^2)$ in $RG^{\text{ret}}_{\bar{x}}$. When $\Delta$ acts on this it yields zero unless $x = (\bar{x}^0, -\bar{x}^1, \bar{x}^2, \bar{x}^3)$. However, this fails to be in $N\backslash\partial N$ if $\bar{x} \in N\backslash\partial N$. □

**Remarks** (3.2.2)

1. Since $R$ reverses the orientation, $\int_M R\omega = -\int_M \omega\, \forall\omega \in E_p$. If supp $j \subset N$ the integral $\int_N (1 + R)dG^{\text{ret}} \wedge J$ can be taken as $\int_M R$ does not send points out of the integration region, and because

$$\int_M R\, dG^{\text{ret}} \wedge J = \int_M R(dG^{\text{ret}} \wedge RJ) = -\int_M dG^{\text{ret}} \wedge RJ$$

the integral can be written as

$$F(\bar{x}) = \int_M dG^{\text{ret}}_{\bar{x}} \wedge (1 - R)J + \int_{\partial M \cap N} \cdots.$$

The components of $-RJ$ are $(-J_0(Rx), J_1(Rx), -J_2(Rx), -J_3(Rx))$, and thus the field produced by $J$ is as if there were a mirror-image charge of reversed sign at $Rx \in CN$, undergoing the reflected motion. It is easy to see that its field taken together with the field directly produced by $J$ satisfies the metallic boundary conditions (3.1.26) on $\partial_v N$ (Figure 36). In actuality there is an induced charge in the metal, not in the interior of $CN$, but rather, according to (3.1.26), on $\partial_v N$. The surface current $\delta(x^1)(i_{dx^1}F)_{|\partial_v N}$ generates the field in $N$ that would come from $-RJ$.

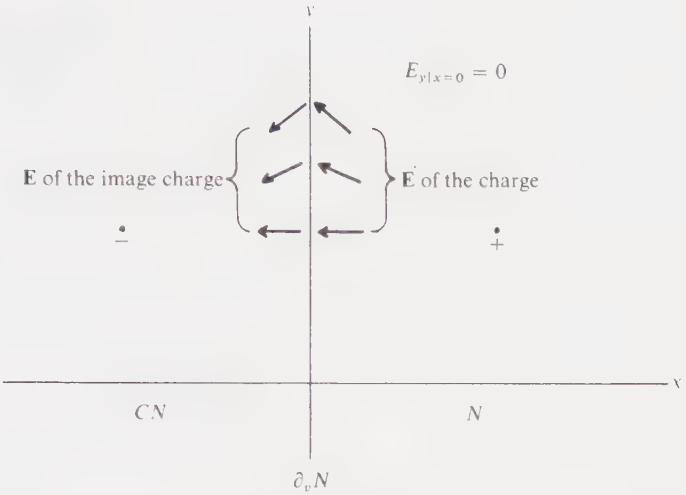



Figure 36 **E** in the presence of a reflecting half-plane

2. It is likewise possible to take $\int_{\partial M \cap N}$ as $\int_{\partial M}$, which involves only $G^{\text{ret}}_{\bar{x}}$, and the appropriate reflected initial data are to be used on $\partial M \cap CN$. If $F$ and ${}^*F_{|\partial M \cap N}$ originate in an incoming wave, then their values in $N$ are the same as if there were no metal present, and the reflected initial values had been specified on $\partial M \cap CN$.
3. The support of $G_x$ in $N$ is contained in the full past light-cone of $\bar{x}$ (Figure 37). Metallic boundary conditions produce echo effects, but never really violate causality.

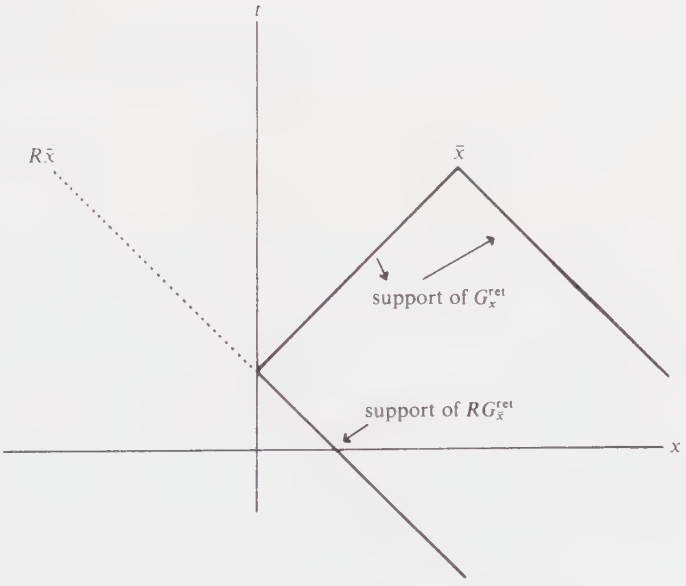



Figure 37 The support of $G_{\bar{x}}$ with a reflecting half-plane

Electromagnetic waves can be transmitted through metallic pipes by a continuous process of reflection. Wave-guides, as these pipes are called, are quite important in communications technology. They are not always made of superconductors, and the limit $\rho e^2/m \to \infty$ is not completely realistic. The field seeps into the walls of the conductor, energy is consumed for Joule heating, and the waves are damped. Nevertheless, the model exhibits the basic mechanism whereby electromagnetic fields can be transmitted like water in a hose.

**The Rectangular Wave-Guide** (3.2.3)

*In the notation of* (3.1.28), *let*

$$M = \{(t, x, y, z) \in \mathbb{R}^4 : t \geq t_0\},$$
$$N = \{x^\alpha \in M : 0 \leq x \leq a, 0 \leq y \leq b\},$$

*and let* $R_1$, $R_2$, $T_1$, *and* $T_2$ *be the transformations on the space of tensors induced by the diffeomorphisms*

$$R_1 : (t, x, y, z) \to (t, -x, y, z)$$
$$R_2 : (t, x, y, z) \to (t, x, -y, z)$$
$$T_1 : (t, x, y, z) \to (t, x + 2a, y, z)$$
$$T_2 : (t, x, y, z) \to (t, x, y + 2b, z)$$

*of* $M$. *Then*

$$G_x = (1 + R_1)(1 + R_2) \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} T_1^n T_2^m G_x^{\text{ret}}$$

*is the Green function of this problem.*

**Proof**

The diffeomorphisms satisfy the commutation relations

$$T_1 T_2 = T_2 T_1, \qquad T_1 R_2 = R_2 T_1, \qquad T_2 R_1 = R_1 T_2,$$
$$R_1 R_2 = R_2 R_1,$$
$$T_1 R_1 = R_1 T_1^{-1}, \qquad T_2 R_2 = R_2 T_2^{-1}.$$

Note that $T_1 R_1$ is the transformation $x \to -x + 2a$, or $(x - a) \to -(x - a)$; in words, a reflection about $x = a$. Similarly, $T_2 R_2$ is a reflection about $y = b$. Since

$$G_{\bar{x}} = R_1 G_{\bar{x}} = R_2 G_{\bar{x}} = T_1 R_1 G_{\bar{x}} = T_2 R_2 G_{\bar{x}}$$

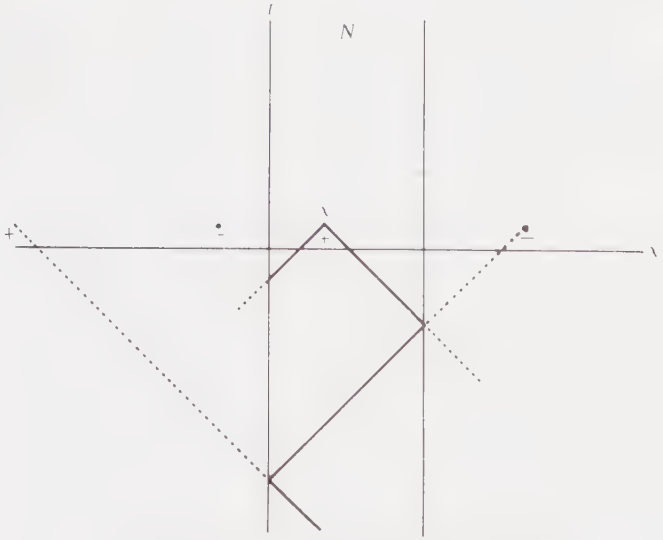



Figure 38 The support of $G_{\bar{x}}$ in a wave-guide

by construction,

$$*dG_{\bar{x}|x=0} = *dG_{\bar{x}|y=0} = *dG_{\bar{x}|x=a} = *dG_{\bar{x}|y=b} = 0,$$

as in the proof of (3.2.1). Moreover, the image of $(0, a)$ under $T_1^n$ equals $(2na, (2n+1)a)$, and its image under $R_1 T_1^n$ equals $(-2na, -(2n+1)a)$, and similar statements hold for $T_2$ and $R_2$. As a result, the only part of the sum $\sum_{n,m}(1+R_1)(1+R_2)T_1^n T_2^m$ that sends a point of $N \backslash \partial N$ back into $N \backslash \partial N$ is the 1 from the term with $n = m = 0$. Hence that is the only term contributing to $\Delta G_{\bar{x}}$ if $x$ and $\bar{x} \in N$, and therefore $-\Delta G_{\bar{x}} = \delta_x$. □

**Remarks** (3.2.4)

1. Although $N$ is not a manifold, because of its sharp edges, its structure is harmless enough that the integration by parts needed in order to use $G_{\bar{x}}$ is easy to justify.
2. Once again, the support of $G_x$ in $N$ is contained in the full past light-cone (Figure 38). This is why we were able to choose $M$ as a region extending to $t = +\infty$ as in (2.2.10).

The structure of $G_{\bar{x}}$ results from the infinite number of reflections of the field back and forth between the walls. Consequently, there are infinitely many image charges, as depicted in cross-section in Figure 39. Their periodic configuration gives rise to characteristic normal modes of oscillation.

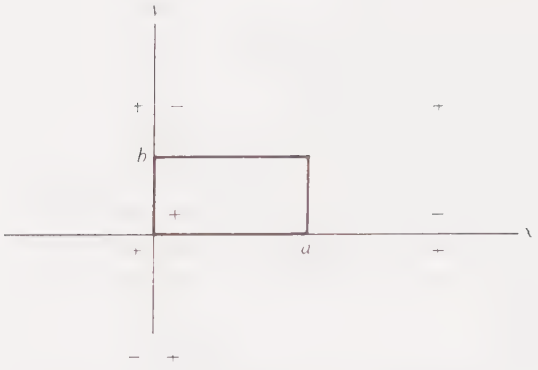



Figure 39 Image charges in a wave-guide

**The Decomposition of $G_x$ into Normal Modes** (3.2.5)

If we write

$$G_{\bar{x}}^{\text{ret}} = \tfrac{1}{2}\bar{e}_{\alpha\beta} \otimes {}^*e^{\alpha\beta} \int \frac{d^4k}{(2\pi)^4} \frac{e^{i\langle k|\bar{x}-x\rangle}}{k^2} \tag{3.2.6}$$

and use the invariance of $e^\alpha$ under the diffeomorphism

$$T_1^n T_2^m : (t, x, y, z) \to (t, x + 2na, y + 2mb, z)$$

then we find ourselves presented with sums of the form

$$\sum_{n=-\infty}^{\infty} \exp(ik(x + 2na) = \exp(ikx) \sum_{g \in \mathbb{Z}} \delta\left(k - \frac{\pi g}{a}\right) \frac{\pi}{a}. \tag{3.2.7}$$

Substitution of this into $G_{\bar{x}}^{\text{ret}}$ and summing $\sum_m$ allows the integration $\int dk_x\, dk_y$ to be accomplished with the aid of delta-functions:

$$\sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \int_{-\infty}^{\infty} dk_x\, dk_y\, e^{i(k_x(x+2na)+k_y(y+2mb))}$$

$$= \frac{\pi^2}{ab} \sum_{g_1=-\infty}^{\infty} \sum_{g_2=-\infty}^{\infty} e^{i\pi(g_1 x/a + g_2 y/b)}. \tag{3.2.8}$$

The $G_{\bar{x}}$ of (3.2.3) becomes

$$G_{\bar{x}} = \tfrac{1}{2}\bar{e}_{\alpha\beta}(1 + R_1)(1 + R_2){}^*e^{\alpha\beta}$$

$$\cdot \sum_{g_1, g_2} \int \frac{d\omega\, dk}{ab(4\pi)^2} \frac{e^{i((\bar{x}-x)\pi g_1/a + (\bar{y}-y)\pi g_2/b + (\bar{z}-z)k - (\bar{t}-t)\omega)}}{(\pi g_1/a)^2 + (\pi g_2/b)^2 + k^2 - \omega^2}. \tag{3.2.9}$$

**The Field Produced by a Current $J$ (3.2.10)**

For the sake of simplicity we consider $M = \mathbb{R}^4$ with no incoming field. We also suppose that the charges are concentrated within $N$, so that $J_{|\partial N} = 0$. Then, after an integration by parts as in (3.1.13),

$$F(\bar{x}) = -\int G_{\bar{x}} \wedge dJ.$$

The Green function $G_{\bar{x}}$ contains the reflections $(1 + R_1)(1 + R_2)$; if we always combine a reflected term with one having a $g_i$ of opposite sign, then, for example for $(1 + R_1)$, we get

$$\int dx(e^{i(\bar{x}-x)\pi g_1/a} \pm e^{-i(\bar{x}+x)\pi g_1/a})dJ(x) = \begin{matrix}\cos \\ i \sin\end{matrix} \bar{x}\pi g_1/a \cdot 2 \int dx\, e^{-ix\pi g_1/a}\, dJ(x).$$

The integration over $\omega$ can be done as in (2.2.25; 2), leaving the explicit formula

$$F = \bar{e}_{\alpha\beta} \sum_{g_1, g_2} f^{\alpha\beta} \int \frac{dk}{2\pi ab} \int d^3x\, dt\, \Theta(\bar{t} - t) \frac{\sin \omega(\bar{t} - t)}{-\omega}$$

$$\cdot e^{ik(\bar{z}-z)} e^{-i(x\pi g_1/a + y\pi g_2/b)}\, dJ^{\alpha\beta}, \tag{3.2.11}$$

where $\omega = ((\pi g_1/a)^2 + (\pi g_2/b)^2 + k^2)^{1/2}$ and $f^{\alpha\beta}$ are the normal modes

$$f = \begin{pmatrix} 0 & ic_1s_2 & is_1c_2 & -s_1s_2 \\ -ic_1s_2 & 0 & c_1c_2 & ic_1s_2 \\ -is_1c_2 & -c_1c_2 & 0 & is_1c_2 \\ s_1s_2 & -ic_1s_2 & -is_1c_2 & 0 \end{pmatrix} \tag{3.2.12}$$

$$c_1 \equiv \cos \bar{x}g_1\pi/a, \qquad s_1 \equiv \sin \bar{x}g_1\pi/a,$$
$$c_2 \equiv \cos \bar{y}g_2\pi/b, \qquad s_2 \equiv \sin \bar{y}g_2\pi/b.$$

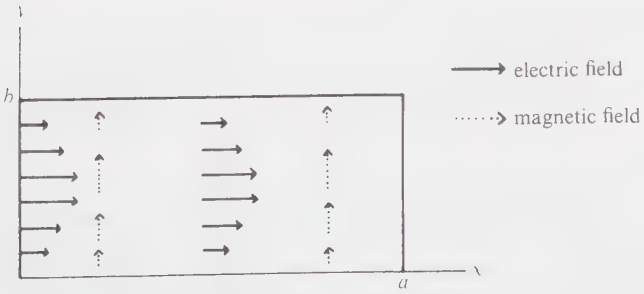



Figure 40 The progression of the fields in a wave-guide

**Remarks** (3.2.13)

1. The boundary conditions (3.1.26) are satisfied because $f_{|\partial N} = 0$. If, say, $\bar{x} = 0$, then they can be checked by deleting the $x^1$ row and column and noting that what is left is $\sim s_1$, which vanishes when $\bar{x} = 0$. It is natural to ask what is meant by the tangential component of **E** or the normal component of **B** at the edges (cf. [2], §7). It turns out that all the components in question are zero at the corners since $s_i = 0$, and only an axial magnetic field $B_3$ is left over.
2. The term with $g_1 = g_2 = 0$ is an axial magnetic field. Electric fields first occur when $g_1 = 0$ and $g_2 = 1$ or vice versa. In the former case the field evolves as shown in Figure 40. The oscillation of the corresponding solution with $J = 0$ (see (3.2.18; 1)) has a minimal frequency $\pi\, b$, and the fields depend on $y$. It is clear that a constant electric field can not exist within a metallic pipe, because the boundary conditions would require it to vanish identically. On the other hand, we know that waves of a sufficiently high frequency are able to travel through metal pipes, because we can see through them.
3. If $g_{1,2} \neq 0$ are fixed, then Equation (3.2.11) describes waves moving in the $z$-direction at a phase velocity $((\pi g_1/ka)^2 + (\pi g_2/kb)^2 + 1)^{1/2} > 1$. It is a purely geometrical effect that this velocity is greater than the speed of light. It comes about because the waves do not pass directly through the pipe, but are reflected back and forth at the walls. A wave moving directly in the $z$-direction could not fulfil the boundary conditions: they demand interference from other waves at a certain angle to the $z$-axis. Such a wave $\sim\exp(i(\mathbf{k}\cdot\mathbf{x} - \omega t))$ with $|\mathbf{k}| = \omega$ necessarily has $\omega/k_z > 1$, as the intersection with a plane of constant phase moves along the $z$-axis faster than light, see Figure 41.
4. The group velocity $\partial\omega/\partial k = k/\omega$ is less than 1. Problem 2 establishes its significance as the flow of energy in the $z$-direction per energy.
5. The question of whether the signal velocity is $\leq 1$ is more pertinent. To answer it, consider the wave-packet

$$g(z, t) = \int_{-\infty}^{\infty} dk\, e^{ikz}[\tilde{g}(k)\cos\omega(k)t + \dot{\tilde{g}}(k)\omega(k)^{-1}\sin\omega(k)t], \quad (3.2.14)$$

where $\tilde{g}$ and $\dot{\tilde{g}}$ are the Fourier transforms of $g$ and $\partial g/\partial t$ at $t = 0$, and

$$\omega(k) = (k^2 + m^2)^{1/2}, \qquad m^2 = \left(\frac{\pi g_1}{a}\right)^2 + \left(\frac{\pi g_2}{b}\right)^2.$$

It is supposed that $g$ is a reasonable function, meaning that it and its first derivatives are integrable, so that all the relevant integrals converge. If $g$ represents a signal, then it must have a well-defined beginning; for instance, $g$ and $\dot{g}$ may be 0 at $t = 0$, $z > 0$. Then the question becomes whether this "wave-front" moves at the speed of light. Indeed it does;

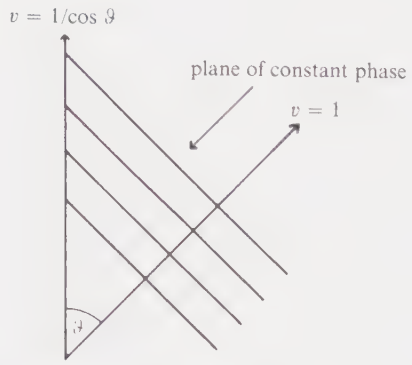



Figure 41 Speeds faster than light in a wave-guide

the above assumptions imply that $g = 0$ for all $z > |t|$ (Problem 5). This does not depend at all on $m^2 > 0$, but holds also if $m^2 < 0$, i.e., for phase velocities less than 1 and group velocities greater than 1 (tachyons) the signal velocity is still 1.

In a wave-guide there are waves with a continuous frequency-spectrum $\omega(k) \geq \omega_{\min}$. If the wave-guide is sealed off in the $z$-direction, then the electromagnetic oscillations have only a discrete spectrum.

**The Resonant Cavity** (3.2.15)

*Let the sets occurring in* (3.1.28) *be*

$$M = \{x^\alpha \in \mathbb{R}^4 : x^0 \geq t_0\},$$
$$N = \{x^\alpha \in M : 0 \leq x^i \leq a_i,\ i = 1, 2, 3\}.$$

*Let $R_i$ denote the reflections $x^i \to -x^i$, and $T_i$ the translations $x^i \to x^i + 2a^i$, as well as the induced transformations of the tensor spaces. Then the Green function for this problem is*

$$G_{\bar{x}} = \prod_{i=1}^{3} \sum_{n_i = -\infty}^{\infty} (1 + R_i) T_i^{n_i} G_{\bar{x}}^{\text{ret}}.$$

The proof proceeds exactly as for (3.2.3), and will not be repeated here. Since the earlier remarks about the causal structure are still valid, let us immediately make the

**Decomposition of $G_{\bar{x}}$ into Normal Modes** (3.2.16)

The formulas of (3.2.5) carry over directly to the case where $N$ is also bounded in the $z$-direction, and after the $\omega$ integration has been done, we find, in analogy to (3.2.9), that

$$G_{\bar{x}} = \tfrac{1}{2}\bar{e}_{\alpha\beta} \otimes \prod_{i=1}^{3} \left(\frac{1 + R_i}{2a_i}\right) * e^{\alpha\beta} \sum_{g_i} e^{i\sum_{j=1}^{3}(\bar{x}^j - x^j)\pi g_j/a_j} \cdot$$

$$\Theta(\bar{t} - t)\frac{1}{\omega}\sin\omega(\bar{t} - t),$$

$$\omega = \left[\sum_{i=1}^{3}\left(\frac{\pi g_i}{a_i}\right)^2\right]^{1/2}.$$

By making a change of sign in the $g_i$ to replace the reflection of the $x_i$ with a reflection of the $\bar{x}_i$, as in (3.2.10), we can solve for the normal modes:

$$G_{\bar{x}} = \tfrac{1}{2}\bar{e}_{\alpha\beta} \otimes \sum_{g_i} f^{\alpha\beta} * e^{\alpha\beta} e^{-i\sum_{j=1}^{3} x^j \pi g_j/a_j} \Theta(\bar{t} - t)\frac{\sin\omega(\bar{t} - t)}{a_1 a_2 a_3 \omega}.$$

This time,

$$f = \begin{pmatrix} 0 & -c_1 s_2 s_3 & -s_1 c_2 s_3 & -s_1 s_2 c_3 \\ c_1 s_2 s_3 & 0 & ic_1 c_2 s_3 & ic_1 s_2 c_3 \\ s_1 c_2 s_3 & -ic_1 c_2 s_3 & 0 & is_1 c_2 c_3 \\ s_1 s_2 c_3 & -ic_1 s_2 c_3 & -is_1 c_2 c_3 & 0 \end{pmatrix},$$

with the abbreviations

$$\left.\begin{matrix} c_i \\ s_i \end{matrix}\right\} = \left\{\begin{matrix} \cos \\ \sin \end{matrix}\right\} \frac{\bar{x}_i g_i \pi}{a_i}.$$

**Remarks** (3.2.17)

1. The boundary conditions are again satisfied at the edges by the disappearance of the relevant components, and at the corners $f$ is identically zero.
2. There are no static fields with $J = 0$ in a resonant cavity; if $g_i = 0$, $i = 1, 2, 3$, then $f \equiv 0$. Consequently there is a minimal frequency

$$\omega_{\min} = \left(\sum\left(\frac{\pi}{a_i}\right)^2\right)^{1/2}.$$

**Problems** (3.2.18)

1. Provide the $f^{\alpha\beta}$ of (3.2.12) with coefficients for which there exists a solution of the homogeneous Maxwell equations ($J = 0$).

2. For the $f$ of Problem 1, calculate the averages of $T^{00}$ and $T^{03}$, and verify that $\overline{T^{03}}/\overline{T^{00}} = k/\omega$.

3. Find the solutions corresponding to Problem 1 for the cylindrical geometry

$$N = \{x^\alpha \in \mathbb{R}^4 : x^2 + y^2 \leq a^2\}.$$

4. What type of oscillation in the cylinder has the lowest $\omega_{\min}$, and how does $\omega_{\min}^2$ times the cross-sectional area compare with the rectangular case?

5. Suppose that the norms

$$\int_{-\infty}^{\infty} dz\, |g(z,0)| \equiv \|g\|, \qquad \int_{-\infty}^{\infty} dz \left| \frac{\partial}{\partial t} g(z,0) \right| \equiv \|\dot{g}\|,$$

and

$$\int_{-\infty}^{\infty} dz \left| \frac{\partial}{\partial z} g(z,0) \right| \equiv \|g'\|$$

are finite for the $g$ of (3.2.14). Show that if $g = \dot{g} = 0$ at $t = 0$, $z > 0$, then $g(z,t) = 0$ for all $z > |t|$.

**Solutions** (3.2.19)

1.

$$f^{\alpha\beta} = \begin{array}{|c|c|c|c|} \hline & i\dfrac{kg_1\pi}{a} c_1 s_2 & i\dfrac{kg_2\pi}{b} s_1 c_2 & (\omega^2 - k^2) s_1 s_2 \\ \hline & & 0 & -i\dfrac{\omega g_1 \pi}{a} c_1 s_2 \\ \hline & & & -i\dfrac{\omega g_2 \pi}{b} s_1 c_2 \\ \hline & & & \\ \hline \end{array} \cdot \exp(i\omega t) \quad \text{or} \quad f \to {}^*f.$$

For this $f$, $B_3 = 0$, and it is known as a $TM$ (transverse magnetic) solution, and hence for ${}^*f$, $E_3 = 0$, which is a $TE$ (transverse electric) solution.

2. Because $\overline{c_i^2} = \overline{s_i^2} = \frac{1}{2}$, etc., we find that

$$\overline{T^{00}} = \frac{\omega^2}{4}(\omega^2 - k^2), \qquad \overline{T^{03}} = \frac{k\omega}{4}(\omega^2 - k^2).$$

3. If $J_n$ denotes the $n$-th Bessel function. then

$$f^{\mu\nu} = \frac{1}{\rho}\left|\begin{array}{cccc} & t & z & \varphi & \rho \\ & & (\omega^2 - k^2)\rho & -\dfrac{kn}{\rho} & ik\rho\dfrac{\partial}{\partial\rho} \\ & (-\omega^2 + k^2)\rho & 0 & -\dfrac{n\omega}{\rho} & i\omega\rho\dfrac{\partial}{\partial\rho} \\ & \dfrac{kn}{\rho} & \dfrac{n\omega}{\rho} & 0 & 0 \\ & -ik\rho\dfrac{\partial}{\partial\rho} & -i\omega\rho\dfrac{\partial}{\partial\rho} & 0 & 0 \end{array}\right| \cdot e^{i(kz+n\varphi-\omega t)} J_n(\rho\sqrt{\omega^2 - k^2}).$$

$$f^{*\mu\nu} = \frac{1}{\rho}\left|\begin{array}{cccc} 0 & 0 & i\omega\dfrac{\partial}{\partial\rho} & n\omega \\ 0 & 0 & ik\dfrac{\partial}{\partial\rho} & kn \\ -i\omega\dfrac{\partial}{\partial\rho} & -ik\dfrac{\partial}{\partial\rho} & 0 & \omega^2 - k^4 \\ -n\omega & -kn & k^2 - \omega^2 & 0 \end{array}\right| \cdot e^{i(kz+n\varphi-\omega t)} J_n(\rho\sqrt{\omega^2 - k^2}).$$

Solving Maxwell's equations for $f_{|\partial\Omega} = 0$ means that the parts enclosed in the dotted lines must vanish when $\rho = a$. This implies that

$$J_n(a\sqrt{\omega^2 - k^2}) = 0, \qquad \text{i.e., } \omega = \pm\sqrt{\frac{k^2 + j_{ni}^2}{a^2}},$$

where $j_{ni}$ is the $i$-th zero of $J_n$. Interchanging $f$ and $^*f$ produces a $TE$ solution, for which it is required that

$$J_n'(a\sqrt{\omega^2 - k^2}) = 0, \qquad \text{where } J_n'(\rho) \equiv d\,J_n(\rho)d\rho,$$

i.e., $\omega_{\min} = j_{ni}'(a)$, where $j_{ni}'$ is the $i$-th zero of $J_n'$.

4. Because $j_{01} = 2.40$, $j_{01}' = 3.83$, and $j_{11}' = 1.84$, the $TE$ solution with $n = 1$ has the lowest $\omega_{\min}$, and its $\omega_{\min}^2 \cdot a^2\pi = \pi(1.84)^2$. This is always somewhat larger than the analogous product for the $g_1 = 1, g_2 = 0$ oscillation with a square cross-section:

$$\left(\frac{\pi}{a}\right)^2 \cdot a^2 = \pi \cdot 3.14 \leq \pi \cdot (1.84)^2 = \pi \cdot 3.38.$$

5. One can analytically continue $g$ into the upper half-plane, where it goes to zero as $|k|^{-1}$, because

$$|\tilde{g}(u+iv)| = \left| \int_{-\infty}^{\infty} dz\, e^{-iz(u+iv)} g(z,0) \right| = \left| \int_{-\infty}^{0} dz\, e^{zv} e^{-izu} g'(z,0) \frac{i}{u+iv} \right|$$

$$\leq \frac{\|g'\|}{\sqrt{u^2+v^2}} \qquad \forall u \in \mathbb{R}, v \in \mathbb{R}^+.$$

Similarly, $\tilde{g}$ is analytic and bounded by $\|g\|$ where $v > 0$. Since $\cos[\omega(k)t]$ and $\sin[\omega(k)t]/\omega(k)$ are entire functions in $k$, the $k$ integral of (3.2.14) can be deformed into the upper half-plane. If we decompose

$$\int_{-\infty}^{\infty} dk \quad \text{into} \quad \int_{\infty}^{-R} dk + \int_{\pi}^{0} i\, d\varphi\, R \exp(i\varphi) + \int_{R}^{\infty} dk,$$

where $k = R\exp(i\varphi)$ is on the semicircle, then $\int_{-\infty}^{-R}$ and $\int_{R}^{\infty}$ go to zero as $R \to \infty$. The remaining integrals are of the form

$$\int_{0}^{\pi} d\varphi \exp[(z \pm t\sqrt{1 + m^2 e^{-2i\varphi} R^2}) R(i\cos\varphi \quad \sin\varphi)] \begin{bmatrix} e^{i\varphi} R\tilde{g}(Re^{i\varphi}) \\ (1 + m^2 e^{-2i\varphi}/R^2)\tilde{g}(Re^{i\varphi}) \end{bmatrix}.$$

Since $|R\tilde{g}(R\exp(i\varphi))|$ and $|\tilde{g}(R\exp(i\varphi))|$ remain bounded as $R \to \infty$, $0 \leq \varphi \leq \pi$, such integrals go to zero if $z > |t|$, by the Riemann–Lebesgue lemma.

## 3.3 Diffraction at a Wedge

*This was the first diffraction problem to succumb to a rigorous treatment. Not only does it confirm the general outlines of one's naive expectations, but it also displays the wave nature of light with a wealth of complex detail.*

The solution of boundary-value problems in somewhat complicated geometrical settings is fraught with difficulties even in the two-dimensional case. For a long time people had to settle for an imprecise theory as handed down primarily from Kirchhoff. The procedure was to substitute into certain exact integral equations the field that would be present if the conductors were removed. By so doing, one obtains a solution of the field equations but violates the boundary conditions. Since the solution involves oscillatory integrals, i.e., the small differences between large positive and negative terms, it is nearly impossible to estimate the errors incurred if one makes a small change in the integrand. Yet the result exhibits the features of experimentally known diffraction patterns, and thus the approximations have remained popular down to the present day among young and old alike. It must nonetheless be counted as a great step forward, that A. Sommerfeld succeeded in solving a nontrivial diffraction problem in 1895; it was only then possible to determine when the approximate theory of diffraction was good, and when it failed.

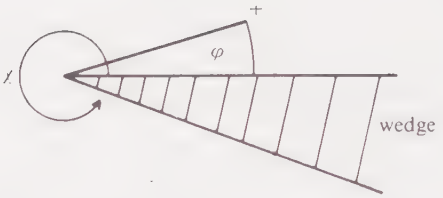



image charges at $-\varphi + 2\chi - 2\pi$

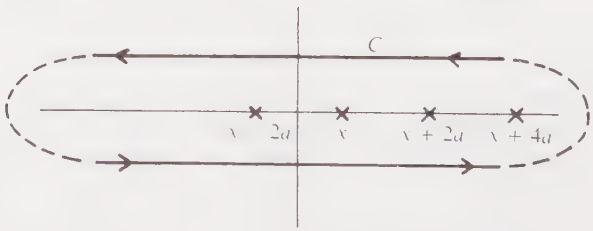
Figure 42 The image charges for a wedge

That problem concerned the diffraction at a wedge, where the metallic surface consisted of the planes where $\varphi = 0$ and where $\varphi = \chi > \pi$. A plausible attempt to solve this problem would proceed by mimicking the solution of the problem of two metallic mirrors at $x = 0$ and $x = a$. Following §3.2, image charges at $x + 2na$ would produce a solution with period $2a$, and then by a reflection at $x = 0$ the boundary conditions on both mirrors could be satisfied. The hitch in trying to transcribe this procedure to the wedge is that if $x$ is the variable, the image charges always remain outside of $0 \le x \le a$ when reflected by $R$, but if $\varphi$ is the variable on the torus $T^1$, then because it is periodic with period $2\pi$, the image charges at $-\varphi + 2n\chi$ would at some point enter the region $0 \le \varphi \le \chi \pmod{2\pi}$. Sommerfeld's brilliant stroke was to forget about the periodicity with period $2\pi$, and to seek a solution with period $2\chi$. His solution has branches, in the sense that by continuation through the metal $\chi \le \varphi \le 2\pi$ the variable would not return to its initial value in $(0, \chi)$, and consequently image charges could appear there. The problem does not require the absence of images at such values of the coordinates; the solution in $0 \le \varphi \le \chi$ has no way of knowing that $\varphi$ is actually a variable on $T^1$.

The most convenient way to construct this ramified solution is to write it as a complex integral. The sums $\sum T^n$ that have appeared above can be represented for analytic $f$ as

$$\sum_n T^n f(x) = \sum_n f(x + 2na) = \int_C \frac{dx'}{2a} \frac{-f(x')}{1 - \exp(2\pi i(x' - x)/2a)}, \tag{3.3.1}$$

where the contour $C$ goes around the poles of the integrand at $x' = x + 2na$.




The path of integration in the integral representation of a sum

The expression (3.3.1) always has a period $2a$ in $x$, as long as the contour $C$ is flexible enough to avoid the poles as $x$ increases to $x + 2a$. The best choice of $C$ depends on the analytic properties of the function $f$.

If we try to apply this procedure to the Green function, we soon run up against the difficulty that $D^{\text{ret}}$ is not an analytic function but only a distribution, $\delta(r - t)\, 4\pi r$. However, it can be approximated by analytic functions, for example by

$$D^{\text{ret}}(x) = \lim_{\varepsilon \downarrow 0} \operatorname{Re} \int_0^\infty d\omega \frac{e^{i(r-t+i\varepsilon)\omega}}{4\pi^2 r}, \tag{3.3.2}$$

where the limit is taken in the sense of distributions. Since the expressions that we require are of the form

$$\int D^{\text{ret}}(\bar{x} - x) f(x) dx,$$

the integrand will be multiplied by the Fourier transform of $f$, and for suitable $f$ the integral $\int d\omega$ will be convergent. The integration becomes rather insensitive to the limit $\varepsilon \downarrow 0$, and therefore we may interchange $\varepsilon \downarrow 0$ with other limiting processes without worry. If we use the above expression and the notation of (3.1.28), we arrive at

**The Green Function for the Wedge** (3.3.3)

*In cylindrical coordinates*

$$x^\alpha = (t, \rho \cos \varphi, \rho \sin \varphi, z),$$

*let*

$$M = \{x^\alpha \in \mathbb{R}^4 : t \geq t_0\}, \qquad N = \{x^\alpha \in M : 0 \leq \varphi \leq \chi\}.$$

*If $R_\varphi$ is the transformation of forms induced by $\varphi \to -\varphi$, then the Green function of this problem is*

$$G_{\bar{x}} = -\tfrac{1}{2} \bar{e}_{\alpha\beta} \otimes (1 + R_\varphi)^* e^{\alpha\beta} \lim_{\varepsilon \uparrow 0} \operatorname{Re} \int_0^\infty d\omega\, e^{-i\omega(\bar{t} - t + i\varepsilon)}$$

$$\int_C \frac{d\varphi'/2\chi}{1 - e^{i\pi(\varphi' - \bar{\varphi} + \varphi)/\chi}} g([(\bar{z} - z)^2 + \bar{\rho}^2 + \rho^2 - 2\bar{\rho}\rho \cos \varphi']^{1/2}),$$

*where*

$$g(r) = \frac{e^{i\omega r}}{4\pi^2 r}.$$

*The integration contour C consists of two curves, one in the upper and one in the lower half-plane, which come in from infinity, enclose the zeros of the term in brackets [ ], avoid the zeroes of the denominator, and then return to infinity (Figure 43).*

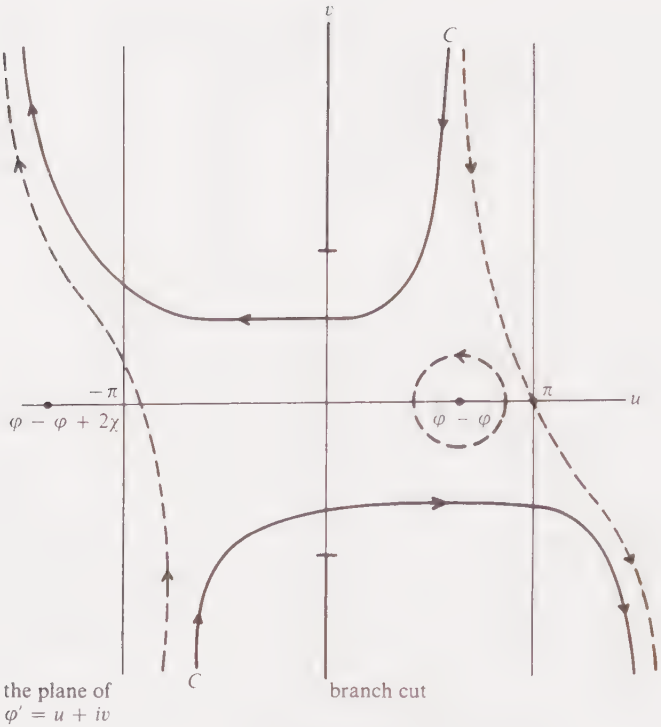



Figure 43 Paths of integration in (3.3.3)

**Proof**

Let $\varphi' = u + iv$, so $\cos\varphi' = \cos u \cosh v - i \sin u \sinh v$. The zeroes of $[\ ] \equiv r$ where $u = 0$ and $\cosh v = ((\bar{z} - z)^2 + \bar{\rho}^2 + \rho^2)/2\bar{\rho}\rho$ require the branch cuts drawn in the figure. Because $\omega > 0$, $\exp(i\omega r)$ decreases exponentially if $\operatorname{Im} r > 0$. If $\varphi'$ is far from the real axis, then $r$ has an imaginary part $\sim \sin u \sinh v$. Therefore $g$ decreases in the upper half-plane if $0 < u < \pi$ or $-2\pi < u < -\pi$, and in the lower half-plane if $-\pi < u < 0$ or $\pi < u < 2\pi$. Since the denominator is small only in the neighborhood of $\varphi' = \bar{\varphi} - \varphi + 2n\chi$, $n \in \mathbb{Z}$, the convergence of the infinite integral is verified if $x \neq \bar{x}$. If $\varphi$ increases by $2\chi$, then the poles move along the real axis, which, however, does not interfere with the integration contour. This immediately implies periodicity in $\varphi$ (and $\bar{\varphi}$) with period $2\chi$, and thus that the boundary conditions are fulfilled exactly as in §3.2.

It remains to verify that $-\Delta G_{\bar{x}} = \delta_{\bar{x}}$. Note that

$$\Delta \frac{e^{i\omega(|\bar{\mathbf{x}} - \mathbf{x}| - \bar{t} + t)}}{|\bar{\mathbf{x}} - \mathbf{x}|} = 0$$

for all $\bar{x} \neq x$, and in particular for complex $\bar{x}$. If we treat $\varphi'$ as a complex part to be added to $\bar{\varphi}$, then we see that the wave equation is satisfied if $x \neq \bar{x}$.

If $x \to \bar{x}$, then the two branch points and the pole move to the origin and pinch the integration contour in two. The integration path is then unable to avoid the point $\bar{\varphi} - \varphi$. In order to see better what the value of the integral is in this limit, let us deform $C$ into the paths drawn with dotted lines in Figure 43 and a circle around $\bar{\varphi} - \varphi$. The integration over the circle just produces $D^{\text{ret}}(\bar{x} - x)$, of course, while the other two parts are not affected by the singularities in the limit $x \to \bar{x}$. This means that $G_x$ differs from $G_{\bar{x}}^{\text{ret}}$ only by a solution of the homogeneous wave equation, and consequently it satisfies everything required of it. □

**Remark** (3.3.4)

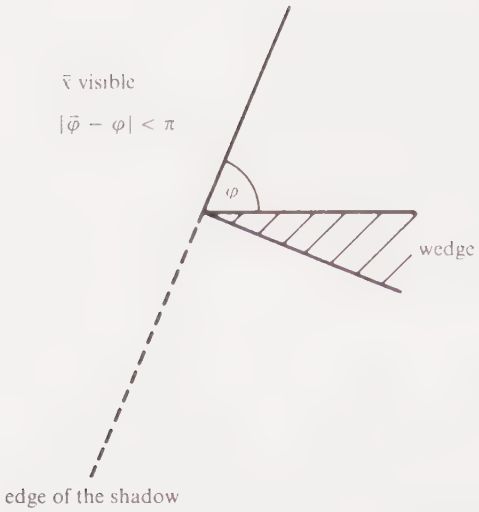
The edge still keeps $N$ from being a manifold. At the edge ($\bar{\rho} \to 0$), $G_{\bar{x}}$ behaves like $(\bar{\rho})^{\pi/\chi}$. Although this approaches zero, if $\chi > \pi$ the derivative, and consequently the nonvanishing components of the field strength, diverge. However, because they approach infinity more slowly than $\rho^{-1}$, the square-integrability is not in danger, and thus the total energy remains finite.

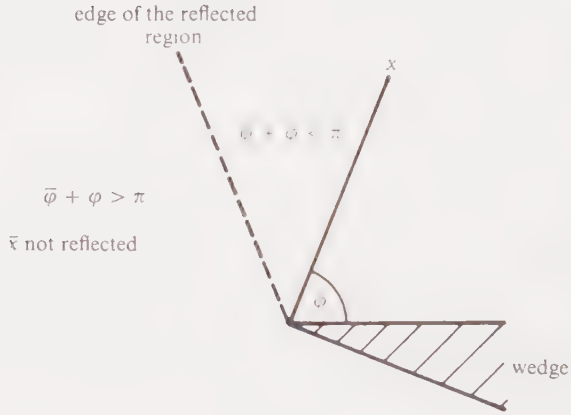
As the first application of (3.3.3) we find out how it accords with the naive idea that the wedge reflects some of the light and casts a shadow. To this end, we assume in the following that $\chi > \pi$.

**Geometric Optics** (3.3.5)

According to geometric optics, if $|\bar{\varphi} - \varphi| < \pi$, then one can see directly from $\bar{x}$ to $x$; if $|\bar{\varphi} + \varphi| < \pi$, then the reflected image from the upper surface is also visible; and if $|\bar{\varphi} - \chi + \varphi - \chi| < \pi$, then the image from the lower surface is visible (Figure 44). These effects can be separated off and the corrections to geometric optics calculated if we make a decomposition of $C$ into the following parts. Two curves, $C_1$ and $C_2$, join $-i\infty - \pi/2$ to $i\infty - 3\pi/2$ and respectively $i\infty + \pi/2$ to $-i\infty + 3\pi/2$, and intersect the real axis at $-\pi$ and respectively $\pi$; and if there is a pole in $(-\pi, \pi)$, then there is also a loop around it (Figure 45).

The loop integral yields $G_{\bar{x}}^{\text{ret}}$, and if $|\bar{\varphi} - \varphi| < \pi$, then we see the light as if there were no wedge, because in the physical region, where $\bar{\varphi}$ and $\varphi \in (0, \chi)$, the poles at $\bar{\varphi} - \varphi + 2n\chi$, $n \neq 0$, are never in $(-\pi, \pi)$. In the reflected part (the one with the $R$ in (3.3.3)), the poles at $\bar{\varphi} + \varphi$ and $\bar{\varphi} + \varphi - 2\chi$ may lie within $(-\pi, \pi)$, which causes light to be reflected by the upper, or respectively lower, surface. Therefore the integrals over $C_1$ and $C_2$ are precisely the corrections to geometric optics. It is apparent that if $\omega\rho \gg 1$ and $\omega\bar{\rho} \gg 1$, that is, at several wavelengths distance from the edge, geometric optics dominates. Because of the factor $\exp(i\omega r)$, the low-lying values of the integrand, in the shaded parts of Figure 45, become deep valleys in this limit. The trail leading up to the pass at $\pm\pi$, where $\exp(i\omega r)$ has absolute value 1, becomes steeper and steeper, and the contribution from going over the top, though dominating


$\bar{x}$ visible
$|\bar{\varphi} - \varphi| < \pi$
$\varphi$
wedge
edge of the shadow


edge of the reflected region
$\bar{x}$
$\bar{\varphi} + \varphi < \pi$
$\bar{\varphi} + \varphi > \pi$
$\bar{x}$ not reflected
$\varphi$
wedge

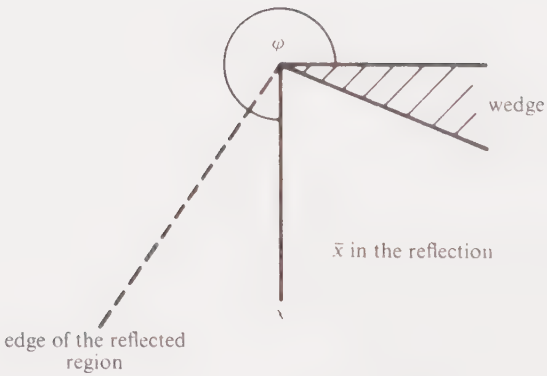

$\varphi$
wedge
$\bar{x}$ in the reflection
edge of the reflected region

Figure 44

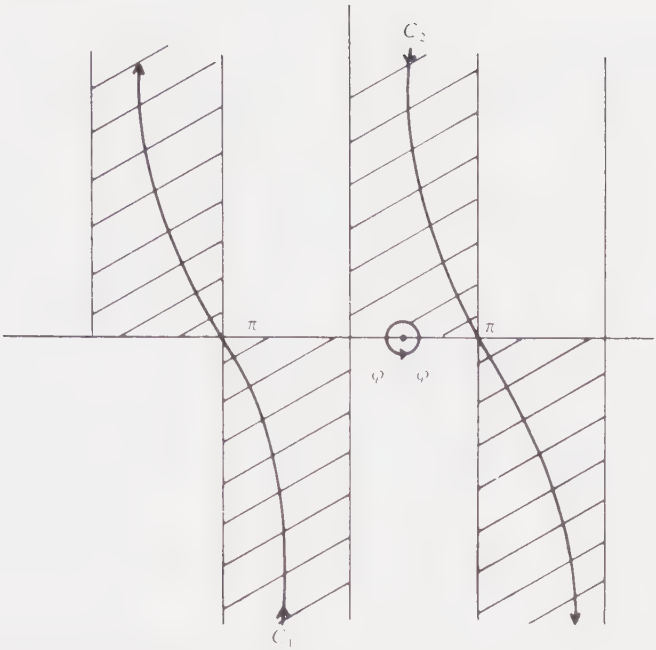



Figure 45 Paths of integration in the hatched regions, where the integrand decreases

the integral, becomes negligible. In order to make these ideas quantitative, let us evaluate

$$\int_{C_1, C_2} \frac{d\varphi'}{2\chi} \frac{e^{i\omega r}}{4\pi r} (1 - e^{i\pi(\varphi' - \bar{\varphi} + \varphi)/\chi})^{-1}$$

by the saddle-point method (= the method of steepest descent). We start by expanding the rapidly varying exponent in the vicinity of the saddle points,

$$r = ((\bar{z} - z)^2 + \bar{\rho}^2 + \rho^2 - 2\bar{\rho}\rho \cos \varphi')^{1/2} \sim R - \frac{\rho\bar{\rho}}{2R} \tilde{\varphi}^2,$$

$$R = ((\bar{z} - z)^2 + (\bar{\rho} + \rho)^2)^{1/2}, \qquad \tilde{\varphi} = \varphi' \pm \pi,$$

and replacing the rest of the integrand by its value at the saddle point. There remains a Gaussian integral over $\tilde{\varphi}$, from which we obtain the asymptotically exact formula:

$$\begin{aligned} &\int_{C_1 \cup C_2} \frac{d\varphi'}{2\chi} \frac{e^{i\omega r}}{4\pi r} (1 - e^{i\pi(\varphi' - \bar{\varphi} + \varphi)/\chi})^{-1} \\ &\quad \cong \frac{1}{4\chi} \left( \frac{1}{1 - e^{i\pi(\bar{\varphi} - \varphi - \pi)/\chi}} - \frac{1}{1 - e^{i\pi(\varphi - \bar{\varphi} + \pi)/\chi}} \right) \cdot \frac{e^{i\omega R - i\pi/4}}{\sqrt{2\pi\omega\bar{\rho}\rho R}} \\ &\quad = -\frac{1}{4\chi} \frac{\sin(\pi^2/\chi)}{\cos \pi(\bar{\varphi} - \varphi)/\chi - \cos \pi^2/\chi} \frac{e^{i\omega R + i\pi/4}}{\sqrt{2\pi\omega\bar{\rho}\rho R}} \end{aligned} \tag{3.3.6}$$

(cf. Problem 1).

**Remarks** (3.3.7)

1. The correction to geometric optics vanishes as $\bar{\rho} \to \infty$. If $\rho \gg \bar{\rho}$, then it goes as $\bar{\rho}^{-1/2}$, which corresponds to a cylindrical wave emanating from the edge.
2. The cylindrical wave has an angular factor, which diverges at the boundary of the shadow, $|\bar{\varphi} - \varphi| = \pi$. However, the asymptotic expansion (3.3.6) does not apply in that limit, because the pole coincides with the saddle point.
3. If $\chi = \pi$, then the cylindrical wave disappears, as the wedge becomes the plane mirror of (3.2.1). This fact is therefore strictly true, and not merely asymptotic (Problem 2).

If $\chi = 2\pi$, then the expression (3.3.3) simplifies so greatly that many additional questions can be answered for which the asymptotic expression (3.3.6) is inadequate.

**Diffraction at a Half-Plane** (3.3.8)

*If $\chi = 2\pi$, $\rho \to \infty$, and $\bar{\rho}$ and $\bar{\varphi}$ are arbitrary, then $G_{\bar{x}}$ is asymptotically*

$$G_{\bar{x}} = (1 + R_\varphi) G_x^{\mathrm{ret}} \frac{e^{-i\pi/4}}{\sqrt{2}} \int_v^\infty dv' \, e^{iv'^2/2},$$

$$v = \sqrt{\frac{\bar{\rho}\rho\omega}{\pi R}}\, 2\cos\left(\frac{\bar{\varphi} - \varphi}{2}\right), \qquad R = ((\bar{z} - z)^2 + \rho^2)^{1/2}.$$

**Remarks** (3.3.9)

1. The $\omega$ in $v$ should be chosen so that $G_\chi^{\mathrm{ret}}$ is represented as the integral of (3.3.2).
2. For (3.3.6) it was necessary to assume that $\omega\bar{\rho}\rho((\bar{z} - z)^2 + \rho^2 + \bar{\rho}^2)^{-1/2} \gg 1$. This is not the case if $\rho \to \infty$ and $\bar{\rho} < 1/\omega$; in this sense (3.3.8) is more general. The boundary of the shadow is also described by (3.3.8), as long as the source of the light is sufficiently far from the edge of the wedge.

**Proof of** (3.3.8)

First observe that

$$G_{\bar{x}}(\varphi) + G_{\bar{x}}(\varphi + 2\pi) = G_{\bar{x}}^{\mathrm{ret}}(\varphi),$$

because

$$\frac{1}{1 - e^{i(\varphi' - \bar{\varphi} + \varphi)/2}} + \frac{1}{1 - e^{i(\varphi' + 2\pi - \bar{\varphi} + \varphi)/2}} = \frac{2}{1 - e^{i(\varphi' - \bar{\varphi} + \varphi)}}.$$

Since this sum has a periodicity $2\pi$, the contributions from the paths $C_1$ and $C_2$ of Figure 45 cancel out, and only the loop around $\bar{\varphi} - \varphi$ remains. On the other hand, the difference of the same terms is

$$\frac{1}{1 - e^{i(\varphi' - \bar{\varphi} + \varphi)/2}} - \frac{1}{1 + e^{i(\varphi' - \bar{\varphi} + \varphi)/2}} = \frac{i}{\sin(\varphi' - \bar{\varphi} + \varphi)/2},$$

and we shall find a similar factor in the exponent in the limit $\rho \to \infty$. In (3.3.3) in this limit, let us write

$$r \cong R - \frac{\bar{\rho}\rho}{R}\cos\varphi', \qquad R = ((\bar{z} - z)^2 + \rho^2)^{1/2},$$

and compare it with the quantity

$$|\bar{\mathbf{x}} - \mathbf{x}| \cong R - \frac{\bar{\rho}\rho}{R}\cos(\bar{\varphi} - \varphi)$$

occurring in $G_x^{\text{ret}}$ in place of $r$: then in the difference there occurs

$$\cos(\bar{\varphi} - \varphi) - \cos\varphi' = 2\cos^2\frac{(\bar{\varphi} - \varphi)}{2} - 2\cos^2\frac{\varphi'^2}{2}$$

$$= 2\sin\frac{(\varphi' + \bar{\varphi} - \varphi)}{2}\sin\frac{(\varphi' - \bar{\varphi} + \varphi)}{2}$$

If the factor $\exp(i\omega|\bar{\mathbf{x}} - \mathbf{x}|)$ is removed, then in the limit $\rho \to \infty$ we are left with the integral

$$I \equiv \frac{1}{8\pi i}\int_C \frac{d\varphi'\, e^{2i\omega(\bar{\rho}\rho/R)\sin(\varphi' - \bar{\varphi} + \varphi)/2\sin(\varphi' + \bar{\varphi} - \varphi)/2}}{\sin(\varphi' - \bar{\varphi} + \varphi)/2}$$

in $G_x(\varphi) - G_{\bar{x}}(\varphi + 2\pi)$. The denominator can be disposed of by differentiating by $\bar{\rho}$;

$$\frac{\partial I}{\partial\bar{\rho}} = \frac{\omega\rho}{4\pi R}\int_C d\varphi' \sin\frac{(\varphi' + \bar{\varphi} - \varphi)}{2}\, e^{(2i\omega\bar{\rho}\rho/R)(\cos^2(\bar{\varphi} - \varphi)/2 - \cos^2\varphi'/2)}.$$

Using

$$\sin\frac{(\varphi' + \bar{\varphi} - \varphi)}{2} = \sin\frac{\varphi'}{2}\cos\frac{\bar{\varphi} - \varphi}{2} + \cos\frac{\varphi'}{2}\sin\frac{\bar{\varphi} - \varphi}{2},$$

we see that the second term contributes nothing, because its part of the integrand is even in $\varphi'$, and the integral $\int_C d\varphi'$ is taken in the opposite direction when $\varphi' \to -\varphi'$. We may choose $\cos\varphi'/2$ as a new integration variable in the first term. Then there is no pole, and the two curves $C_1$ and $C_2$ have the same contribution, since the integral changes its sign under $\varphi' \to \varphi' + 2\pi$:

$$\frac{\partial I}{\partial\bar{\rho}} = \frac{\omega\rho}{\pi R}\int_{-i\infty}^{i\infty} e^{-2i(\omega\bar{\rho}\rho/R)(t^2 - \cos^2(\bar{\varphi} - \varphi)/2)}\, dt\cos\frac{\bar{\varphi} - \varphi}{2}$$

$$= \frac{\omega\rho}{\pi R}e^{-i\pi/4}e^{2i(\omega\bar{\rho}\rho/R)\cos^2(\bar{\varphi} - \varphi)/2}\sqrt{\frac{2\pi R}{\omega\bar{\rho}\rho}}\cos\frac{\bar{\varphi} - \varphi}{2}.$$

In order to do the $\bar{\rho}$ integration, we introduce the variable

$$v = \sqrt{\frac{\bar{\rho}\rho\omega}{\pi R}}\, 2\cos\frac{(\bar{\varphi} - \varphi)}{2},$$

$$dv = d\bar{\rho}\sqrt{\frac{\rho\omega}{\pi\bar{\rho}R}}\cos\frac{(\bar{\varphi} - \varphi)}{2}.$$

Since $I = 0$ for $\bar{\rho} = 0$ where $C$ can be shifted away

$$I = \sqrt{2}e^{-i\pi/4}\int_0^v dv'\, e^{i\pi v'^2/2}.$$

Combining this with the sum $G_{\bar{x}}(\varphi) + G_x(\varphi + 2\pi)$ and using

$$\sqrt{2}e^{-i\pi/4}\int_{-\infty}^0 dv'\, e^{i\pi v'^2/2} = 1,$$

we obtain (3.3.8). □

To discuss the problem in more detail, we need some

**Properties of Fresnel's Integral** (3.3.10)

Let

$$F(z) = \int_{-\infty}^z dv\, e^{i\pi v^2/2}.$$

Then $F(\infty) = 2F(0) = 1 + i$, and $F$ behaves asymptotically as

$$\begin{aligned} z \to 0&: F(z) = F(0) + z + i\frac{\pi}{6}z^3 + O(z^5) \\ z \to \infty&: F(z) = F(\infty) + \frac{e^{i\pi z^2/2}}{i\pi z} + O(z^{-3}) \\ z \to -\infty&: F(z) = \frac{e^{i\pi z^2/2}}{i\pi z} + O(z^{-3}). \end{aligned} \tag{3.3.11}$$

If $F(x)$ is graphed as a curve in the complex plane depending on $x \in \mathbb{R}$ as a parameter, it has the form of Cornu's spiral. When $x \le x_0 > 0$, $|F(x)|$

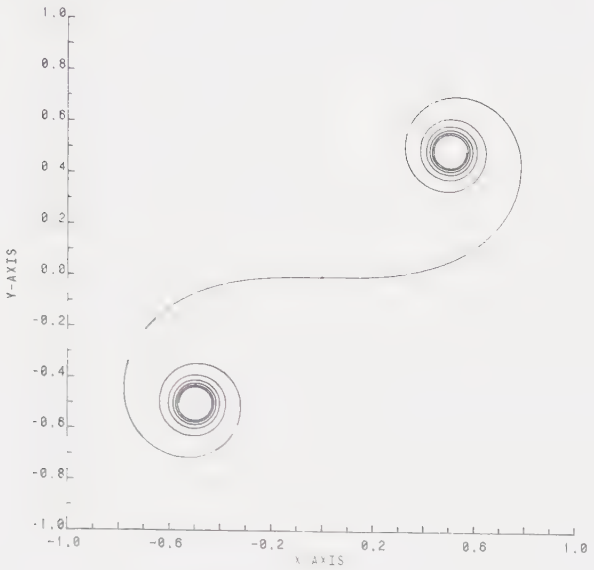



Figure 46 Cornu's spiral

increases monotonically, and after that point it oscillates about the limiting value $\sqrt{2}$.

We next investigate the regions not covered by (3.3.6):

**The Boundary of the Shadow** (3.3.12)

Let $\bar{\varphi} - \varphi = \pi + \delta$, $\delta \ll 1$. The time average of the intensity equals the absolute square of the field, and its ratio to the unscreened intensity $i^{\text{ret}}$ becomes

$$\frac{i}{i^{\text{ret}}} = \frac{1}{2}\left|F\left(\delta\sqrt{\frac{\bar{\rho}\omega}{\pi}}\right)\right|^2$$

as $\rho \to \infty$ with $\rho/R \to 1$. We have disregarded the reflected light; the interference with the term containing $R\varphi$ produces a correction $O(\bar{\rho}^{-1/2})$. During the transition from light to shadow, the intensity thus first oscillates about the value $i^{\text{ret}}$ and then decreases monotonically. The width $\sim\sqrt{\bar{\rho}/\omega}$ of the transition zone, where the shadow is hazy, increases with $\bar{\rho}$, while its angle $1/\sqrt{\bar{\rho}\omega}$ from the edge of the wedge approaches zero as $\bar{\rho}$ increases.

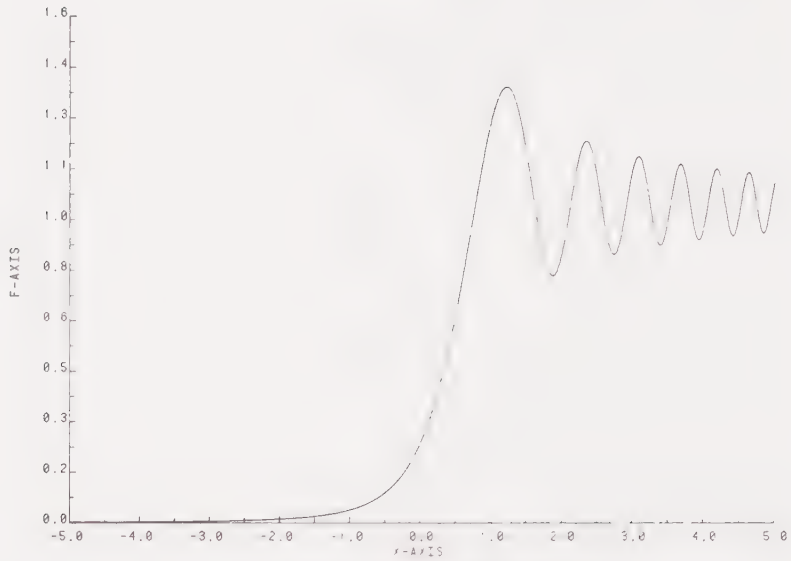



Figure 47 The intensity at the edge of a shadow

**The Edge** (3.3.13)

If $\bar{\rho} \to 0$ and $\rho/R \to 1$, and we consider such a small neighborhood of the edge that the contribution from $G_x^{\text{ret}}$ is effectively constant and hence unaffected by a reflection, then the components that go to zero at the edge are

$$F\left(\sqrt{\frac{\bar{\rho}\omega}{\pi}}\,2\cos\left(\frac{\bar{\varphi}-\varphi}{2}\right)\right) - F\left(\sqrt{\frac{\bar{\rho}\omega}{\pi}}\,2\cos\left(\frac{\bar{\varphi}+\varphi}{2}\right)\right) \sim \sqrt{\frac{\bar{\rho}\omega}{2}}\,2\sin\frac{\bar{\varphi}}{2}\cdot\sin\frac{\varphi}{2}.$$

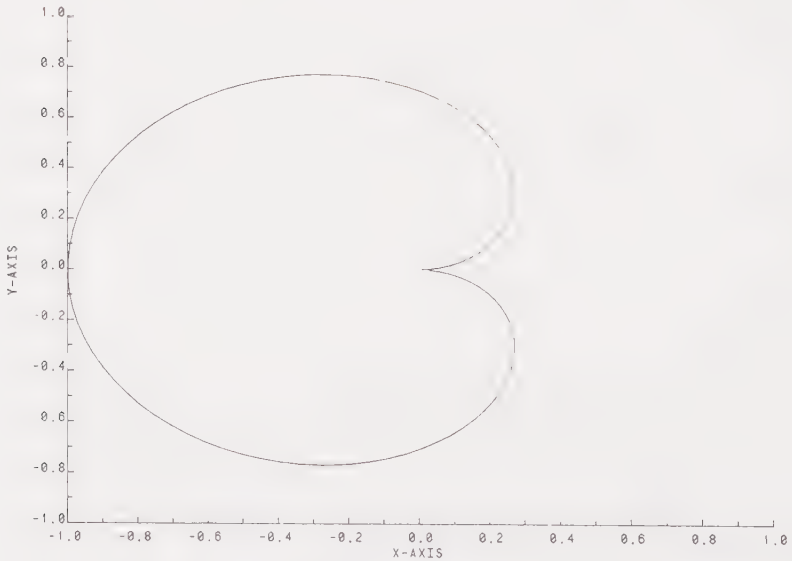



Figure 48 Polar graph of $i(\varphi)$

Thus the intensity dies down in directions along the half-plane, and the angular distribution in $\bar{\varphi}$ with $\varphi$ fixed has no preference for the incident direction.

**The Plane of the Screen** (3.3.14)

If $\bar{\varphi} = \pi$, and the factor $G_{\bar{x}}^{\text{ret}}$ is still effectively constant, then the components that vanish at the screen are diminished by

$$F\left(2\sqrt{\frac{\bar{\rho}\omega}{\pi}}\sin\frac{\varphi}{2}\right) - F\left(-2\sqrt{\frac{\bar{\rho}\omega}{\pi}}\sin\frac{\varphi}{2}\right)$$

compared with $G_{\bar{x}}^{\text{ret}}$. It takes a few wavelengths for the field to increase from zero at $\bar{\rho} = 0$ to the value it would have were it not for the screening, and it takes longer to do so if the light is incident at a smaller angle $\varphi$

**Remark** (3.3.15)

In the integral, Kirchhoff's theory of diffraction replaces the field within the screening plane with the value it would have without screening. This is a very good approximation at many wavelengths away from the screen, but it is poor in the vicinity of the screen. It has the unsettling consequence that the field strength calculated in this manner becomes singular at the edge as $1/\rho$, and is thus not square-integrable, so the field energy is infinite (cf. [2], §25).

In conclusion, geometric optics gives the right answers if one can take a bird's-eye view, from which the wavelength is not perceptible. But up close there is such a complicated pattern of interference that any simple approximation is doomed to failure.

**Problems** (3.3.16)

1. Show how and where (3.3.6) is an asymptotic expansion.
2. Check that $G_{\bar{x}} = (1 + R)G_{\bar{x}}^{\text{ret}}$ if $\chi = \pi$.
3. See whether (3.3.6) agrees with (3.3.8).
4. How large is the error in (3.3.11)?

**Solutions** (3.3.17)

1. Use the saddle-point method discussed, e.g., in Dieudonné, *Infinitesimal Calculus* IX, §1.
2. If $\chi = \pi$, then the function of period $2\chi$ reduces to the function $G_{\bar{x}}^{\text{ret}}$ of period $2\pi$.
3. Applying (3.3.11) to (3.3.8), we get

$$\frac{e^{i\omega|\bar{x}-x|}}{4\pi|\bar{x}-x|}\frac{e^{-i\pi/4}}{\sqrt{2}}\frac{e^{i\pi v^2/2}}{i\pi v} \cong -\frac{e^{i\omega R}ie^{-i\pi/4}e^{i\omega\bar{\rho}\rho/R}}{8\pi\sqrt{2\pi\omega\bar{\rho}\rho R\cos(\bar{\varphi}-\varphi)/2}}$$

for the correction to geometric optics, in agreement with (3.3.6) (where the $R$ is $\sim R + \bar{\rho}\rho/R$).

4. By integration by parts with $z \in \mathbb{R}^+$,

$$-\int_z^\infty dv\, e^{i\pi v^2/2} = -\int_z^\infty \frac{dv}{i\pi v}\frac{\partial}{\partial v} e^{i\pi v^2/2} = \frac{e^{i\pi v^2/2}}{i\pi v} - \int_z^\infty \frac{dv}{i\pi v^2} e^{i\pi v^2/2},$$

$$\left|\int_z^\infty \frac{dv}{i\pi v^2} e^{i\pi v^2/2}\right| = \left|\int_z^\infty \frac{dv}{i\pi^2 v^3}\frac{\partial}{\partial v} e^{i\pi v^2}\right| \le \frac{1}{\pi^2 z^3} + \frac{1}{\pi^2 z^3}.$$

Therefore

$$\left|F(z) - F(\infty) - \frac{e^{i\pi z^2/2}}{i\pi z}\right| \le \frac{2}{\pi^2 z^3} \qquad \forall z \in \mathbb{R}^+.$$

## 3.4 Diffraction at a Cylinder

*This problem is so tractable mathematically that all the various phenomena of geometric and wave optics can be worked out.*

Despite their being complementary problems to the wave-guide and the resonant cavity, it is an elaborate job to analyze diffraction at a cylinder or at a sphere. Even if it is not required to write down the complete Green function, but is enough to solve for the individual waves that $G_\lambda$ is constructed from, there are still infinite sums and complex integrals to cope with. These expressions reduce to elementary functions only in particular limits, in which the relevant properties are well displayed. Depending on the ratio of the wavelength to the size of the metallic body, it may be possible for the wave to pass around the obstruction. This phenomenon is somewhat simpler for a cylinder than for a sphere, for which reason we restrict our discussion to cylinders. We also consider only scalar solutions $u$ to the wave equation $-\Box u = 0$, with either Dirichlet or Neumann boundary conditions, $u_{|\partial N} = 0$ or $du_{|\partial N} = 0$. Although it is possible to construct $G_\lambda$ with the aid of these solutions, the expressions that result are too involved to be illuminating. We are again primarily interested in the limit where the source of the waves is at infinity; therefore we utilize $u$'s incident in a plane and scattering into cylindrical waves. In cylindrical coordinates they would have the form

$$\left(e^{-iqx} + \frac{e^{iq\rho}}{\sqrt{\rho}} f(\varphi)\right) e^{i(kz - \omega t)}$$

as $\rho \to \infty$. If the first term is expanded in a Fourier series,

$$e^{iqx} = \sum_{n=-\infty}^{\infty} e^{in\varphi} \frac{1}{2\pi}\int_0^{2\pi} d\varphi'\, e^{-i(q\rho\cos\varphi' + n\varphi')} = \sum_{n=-\infty}^{\infty} e^{in(\varphi - \pi/2)} \mathrm{J}_n(q\rho), \quad (3.4.1)$$

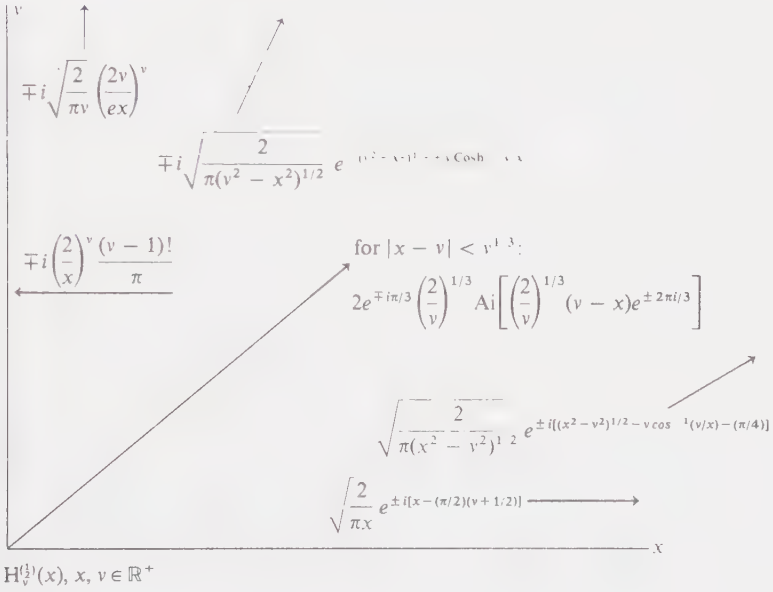
then we can represent the solution $u(\rho, \varphi)$ that satisfies the boundary conditions at $\rho = a$ and $\rho \to \infty$ as a

**Fourier Series** (3.4.2)

$$u = \sum_{n=-\infty}^{\infty} e^{in(\varphi - \pi/2)} \left[ J_n(q\rho) - \frac{J_n(qa)}{H_n^{(1)}(qa)} H_n^{(1)}(q\rho) \right].$$

In order to evaluate the sum, recall the

**Asymptotic Behavior of** $H_\nu^{(1)}(x)$ (3.4.3)


$\nu$
$\mp i\sqrt{\frac{2}{\pi\nu}}\left(\frac{2\nu}{ex}\right)^{\nu}$
$\mp i\sqrt{\frac{2}{\pi(\nu^2 - x^2)^{1/2}}}\, e^{[(\nu^2 - x^2)^{1/2} - \nu \cosh^{-1}(\nu/x)]}$
$\mp i\left(\frac{2}{x}\right)^{\nu}\frac{(\nu-1)!}{\pi}$
for $|x - \nu| < \nu^{1/3}$:
$2e^{\mp i\pi/3}\left(\frac{2}{\nu}\right)^{1/3} \mathrm{Ai}\left[\left(\frac{2}{\nu}\right)^{1/3}(\nu - x)e^{\pm 2\pi i/3}\right]$
$\sqrt{\frac{2}{\pi(x^2 - \nu^2)^{1/2}}}\, e^{\pm i[(x^2-\nu^2)^{1/2} - \nu\cos^{-1}(\nu/x) - (\pi/4)]}$
$\sqrt{\frac{2}{\pi x}}\, e^{\pm i[x - (\pi/2)(\nu + 1/2)]}$
$x$
$H_\nu^{(1)}(x),\ x, \nu \in \mathbb{R}^+$

**Remarks** (3.4.4)

1. The asymptotic behavior of the other solution follows from

$$H_\nu^{(2)}(x) = (H_\nu^{(1)}(x^*))^*, \qquad H_{-\nu}^{(1)}(x) = e^{\pm i\pi\nu} H_\nu^{(1)}(x),$$
$$J_\nu = \tfrac{1}{2}(H_\nu^{(1)} + H_\nu^{(2)}),$$
$$J_\nu \overset{\nu \to \infty}{\sim} \frac{1}{\sqrt{2\pi\nu}} \left(\frac{ex}{2\nu}\right)^{\nu}, \qquad \overset{x \to 0}{\sim} \left(\frac{x}{2}\right)^{\nu} \frac{1}{\nu!}.$$

2. Consequently, if $n$ is large, then the $n$-th term in (3.4.2) approaches

$$\frac{e^{in(\varphi-\pi/2)}}{\sqrt{2\pi n}}\left(\frac{e}{2n}\right)^n\left[(q\rho)^n-\frac{(qa)^{2n}}{(q\rho)^n}\right],$$

and the sum converges uniformly on compact subsets of the $x$-$y$-plane. Even so, the first term in the sum dominates the scattered wave only if $qa \ll 1$. In that case,

$$\frac{J_n(qa)}{H_n^{(1)}(qa)} \to \pi i\left(\left(\frac{qa}{2}\right)^{|n|}\frac{1}{|n|!}\right)^2 n \quad \text{for } |n| \geq 1,$$

and

$$u = e^{-iqx} + \frac{i\pi}{2\ln qa}H_0^{(1)}(q\rho) + O((qa)^2).$$

3. If $qa \gg 1$, then we make use of the fact that $H_\nu^{(1)}$ are analytic functions in $\nu$. This allows the sum in (3.4.2) to be turned into an integral over $d\nu$,

$$\sum_{n=\cdot}^{\infty} = \frac{1}{2i}\int_C \frac{d\nu\, e^{-i\pi\nu}}{\sin \pi\nu}, \tag{3.4.5}$$

where $C$ is a contour passing around the integers in the complex plane and excluding the zeros of $H_\nu^{(1)}$ (Figure 49), which we shall investigate later.

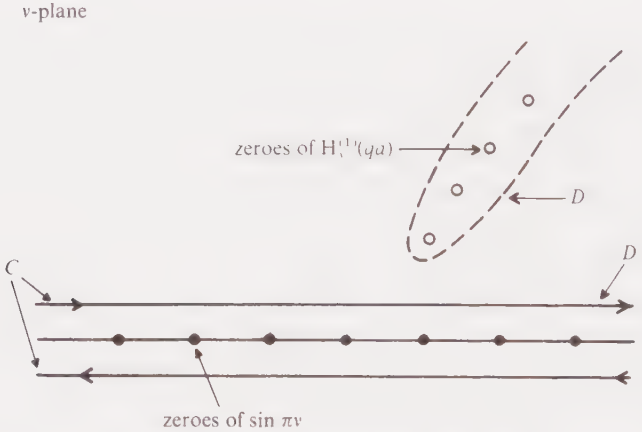



Figure 49 Paths of integration for (3.4.6)

This integral can be further transformed into an integral along the path $D$ strictly above the real axis, by taking account of (3.4.4; 1). The result of this is the

**Representation as a Fourier Integral** (3.4.6)

$$u = \frac{1}{4i}\int_D \frac{d\nu\, e^{-i\nu\pi/2}}{\sin \pi\nu} 2\cos\nu(\varphi-\pi)$$
$$\cdot\frac{H_\nu^{(1)}(qa)H_\nu^{(2)}(q\rho) - H_\nu^{(2)}(qa)H_\nu^{(1)}(q\rho)}{H_\nu^{(1)}(qa)}.$$

To evaluate the above integral by means of the residue theorem, we need to also know the

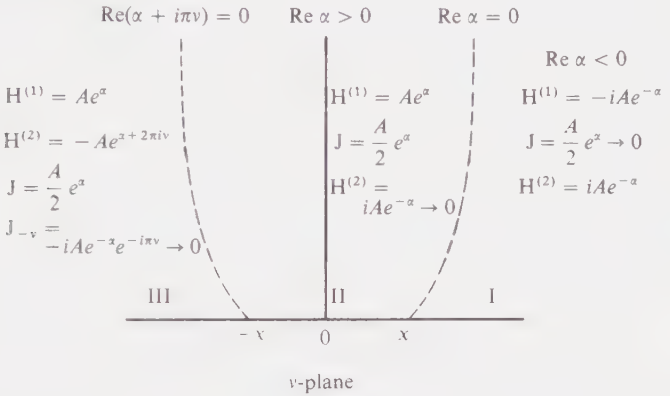
**Asymptotic Behavior of the Bessel Functions in the Upper ν-Plane** (3.4.7)

Suppose that $x$ is real, and define

$$A \equiv \sqrt{\frac{2}{\pi}}(\nu^2 - x^2)^{-1/4} \to \sqrt{\frac{2}{\pi\nu}},$$

$$e^{\alpha} \equiv \exp(\nu^2 - x^2)^{1/2}\left[\frac{\nu}{x} + \frac{\sqrt{\nu^2 - x^2}}{x}\right]^{-\nu} \to \left(\frac{ex}{2\nu}\right)^{\nu}.$$

Then for large $x$ and $|\nu|$, $H^{(\frac{1}{2})}$ and J behave as:




**Remarks** (3.4.8)

1. The function J is always of the form $(A/2)e^{\alpha}$, whereas $H^{(1)}$ (respectively $H^{(2)}$) changes its asymptotic behavior along the curve $\operatorname{Re}\alpha = 0$ (respectively $\operatorname{Re}(\alpha + i\pi\nu) = 0$), on which its zeroes are situated. These curves intersect the real axis at $\pm x$ at an angle of $\pm\pi/3$. Far away from the curve $\operatorname{Re}\gamma = 0$, $H^{(1)}$ increases rapidly in the upper half-plane, going as $|\nu|^{\operatorname{Re}\nu}$.
2. The zeroes $\nu_m$, $m = 1, 2, 3, \ldots$, of $H_{\nu}^{(1)}(qa)$ closest to the real axis are near $\nu = qa$, where the asymptotic representation in terms of Airy functions is applicable; if $qa \gg 1$, then

$$H_{\nu_m}^{(1)}(qa) = 0, \qquad \nu_m \cong qa + c_m\left(\frac{qa}{2}\right)^{1/3} e^{i\pi/3},$$

$$\operatorname{Ai}(-c_m) = 0, \qquad c_m = (2.3, 4.1, 5.5, 6.7, 7.9, \ldots).$$

3. Farther from the real axis along the curve $\operatorname{Re}\alpha = 0$, $H^{(1)}$ is given simply as the sum of the asymptotic forms of regions I and II, and becomes zero if

$$\operatorname{Sinh}\left(\alpha - \frac{i\pi}{4}\right) = 0.$$

This implies that

$$\nu_m \sim \frac{i\pi m}{\ln \dfrac{2\pi m}{eqa}} e^{-(i\pi/2)\ln(2\pi m/eqa)}, \qquad m \gg 1.$$

These facts allow us to discover expressions for $u$ in different regions of the $\rho$-$\varphi$-plane when $qa \gg 1$.

**The Shadow** (3.4.9)

The integral in (3.4.6) can be turned into an integral over $D'$ around the zeroes of $H_\nu^{(1)}(qa)$, because the integrand goes to zero exponentially fast in the upper half-plane, except along the curve $\operatorname{Re}\nu = 0$. To see the reason for this, consider the numerator

$$g_\nu \equiv H_\nu^{(1)}(qa)H_\nu^{(2)}(q\rho) - H_\nu^{(2)}(qa)H_\nu^{(1)}(q\rho) = g_{-\nu}. \tag{3.4.10}$$

In the region $I \cup II$ of (3.4.7), within which the asymptotic form of $H_\nu^{(1)}$ changes, we express $H_\nu^{(1)}$ in terms of J and $H_\nu^{(2)}$. Meanwhile, for the same reason we write $H_\nu^{(2)}$ in terms of $J_{-\nu}$ in $II \cup III$:

$$\begin{aligned} g_\nu &= 2[J_\nu(qa)H_\nu^{(2)}(q\rho) - H_\nu^{(2)}(qa)J_\nu(q\rho)] \\ &= 2e^{i\pi\nu}[H_\nu^{(1)}(qa)J_{-\nu}(q\rho) - J_{-\nu}(qa)H_\nu^{(1)}(q\rho)]. \end{aligned} \tag{3.4.11}$$

In either case (3.4.7) shows us the asymptotic expression

$$g_\nu \sim \frac{2i}{\pi\nu}\left[\left(\frac{a}{\rho}\right)^\nu - \left(\frac{\rho}{a}\right)^\nu\right], \tag{3.4.12}$$

which thus holds uniformly in the upper half-plane. Hence $g_\nu$ is essentially $c^{\pm\nu}$, whereas the denominator $H_\nu^{(1)}(qa)$ is asymptotically dominated by $\nu^{\pm\nu}$, which wins out over $c^{\pm\nu}$, as can be seen if we let $\nu = t\exp(i\psi)$, $t$ and $\psi \in \mathbb{R}^+$; then for $t \to \infty$,

$$\begin{aligned} |\nu^{\pm\nu}| &= e^{t|\cos\psi \ln t - \psi\sin\psi|}, \\ |c^{\pm\nu}| &= e^{t|\cos\psi \ln c|}. \end{aligned} \tag{3.4.13}$$

Because all the other factors in (3.4.6) grow at the fastest as $c^{\pm\nu}$, the integrand decreases exponentially in the upper half-plane for all $\rho$ and $\varphi$. An exception is made, of course, for the curve $\operatorname{Re}\nu = 0$ (asymptotically, $\ln t = \psi\tan\psi$), because the denominator can vanish, and the integral $\int_{D'}$ can be rewritten as the sum of the residues:

$$u = \pi\sum_{m=1}^{\infty} \frac{e^{-i\nu_m\pi/2}}{\sin\pi\nu_m}\cos(\varphi - \pi)\nu_m \frac{H_{\nu_m}^{(2)}(qa)}{\dfrac{\partial}{\partial\nu_m}H_{\nu_m}^{(1)}(qa)}H_{\nu_m}^{(1)}(q\rho). \tag{3.4.14}$$

This equation is only useful if the term with $m = 1$ is so much larger than the rest of the sum that it is essentially equal to $u$. To discover when this happens, we refer again to (3.4.3) to calculate

$$r_m \equiv \frac{H_{\nu_m}^{(2)}(qa)}{\frac{\partial}{\partial \nu_m} H_{\nu_m}^{(1)}(qa)} \tag{3.4.15}$$

for small $m$, with $\nu_m$ as in (3.4.8; 2). We learn that the order of magnitude of this factor is the same for the first few $m$'s.

If $\rho - a \gg a(qa)^{-2/3}$, then the other asymptotic form can be used for $H_{\nu_m}^{(1)}(q\rho)$,

$$H_{\nu_m}^{(1)}(q\rho) = \sqrt{\frac{2}{\pi}}(q^2\rho^2 - \nu_m^2)^{-1/4} \cdot \exp\left\{i\left[(q^2\rho^2 - \nu_m^2)^{1/2} - \nu_m \arccos\frac{\nu_m}{q\rho} - \frac{\pi}{4}\right]\right\}. \tag{3.4.16}$$

Whether the summands of (3.4.14) decrease rapidly with $m$ depends mainly on whether the coefficient of $i\nu_m$ in the exponent is positive. Since

$$\operatorname{Im} \nu_m = \left(\frac{qa}{2}\right)^{1/3} c_m \sqrt{\frac{2}{3}} \gg 1,$$

we see that

$$\sin \pi\nu_m \cong \frac{e^{-i\pi\nu_m}}{2i};$$

and if $|\varphi - \pi| \gg (qa)^{-1/3}$, then also

$$\cos(\varphi - \pi)\nu_m \cong \tfrac{1}{2} e^{-i\nu_m|\varphi - \pi|}.$$

All together, there is an overall factor (up to terms $O((qa)^{-1/2})$) of

$$\exp\left\{i\nu_m\left[\frac{\pi}{2} - |\varphi - \pi| - \arccos\frac{a}{\rho}\right]\right\},$$

making the condition for the first term of (3.4.14) to dominate that

$$|\varphi - \pi| + \arccos\frac{a}{\rho} - \frac{\pi}{2} < -(qa)^{-1/3}. \tag{3.4.17}$$

This amounts to the geometric condition that there is a shadow with a boundary fuzzy within an angle $\sim(qa)^{-1/3}$ (Figure 50). With the numerical

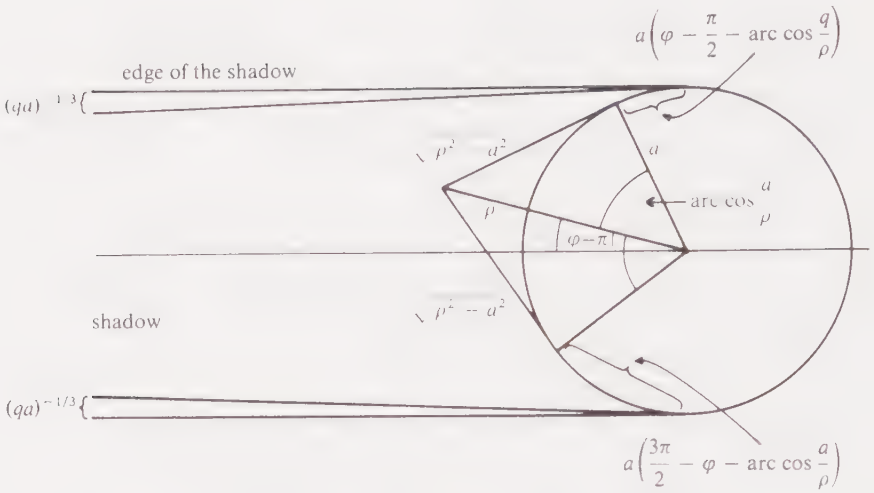



Figure 50 The shadow according to geometric optics

value of $c_1$, the resulting expression for $u$, up to corrections $\sim (qa)^{-1/3}$, is

$$u \cong \sqrt{\frac{\pi}{2q}} \exp\left(\frac{i\pi}{4}\right) r_1 \frac{\exp(iq\sqrt{\rho^2 - a^2})}{(\rho^2 - a^2)^{1/4}}$$
$$\cdot \left[\exp(iqa - 1.5(qa)^{1/3})\left(\varphi - \frac{\pi}{2} - \arccos\frac{a}{\rho}\right)\right.$$
$$\left. + \exp(iqa - 1.5(qa)^{1/3})\left(-\varphi + \frac{3\pi}{2} - \arccos\frac{a}{\rho}\right)\right]. \quad (3.4.18)$$

The interpretation of this is that waves impinging on $(x, y) = (0, \pm a)$ are initially attenuated as they propagate along the surface to the points from which they can emanate in a straight line to the reference position. The damping factor of such "creeping waves" is correlated with the fuzziness of the shadow. During the linear propagation through the distance $\sqrt{\rho^2 - a^2}$ there is the usual amplitude factor $(\rho^2 - a^2)^{-1/4}$ for cylindrical waves, along with the phase factor $\exp(iq\sqrt{\rho^2 - a^2})$.

**The Illuminated Region at $|\varphi| > \pi/2$ (3.4.19)**

Since (3.4.14) is no good outside the shadow, we rewrite the $\cos \nu(\varphi - \pi)$ of (3.4.6) as $\exp(i\nu\pi)\cos \nu\varphi - i \exp(i\nu\varphi)\sin \nu\pi$. Taking the first term of this expression changes (3.4.17) to

$$|\varphi| + \arccos\frac{a}{\rho} \leq \frac{3\pi}{2} - (qa)^{-1/3}, \quad (3.4.20)$$

and this is always true except exactly in the forward direction. Hence this part produces only the damped waves that manage to slip by the cylinder, and can be neglected in comparison with the remainder, $u_2$. The $\sin \nu\pi$ cancels out of $u_2$, so the integration contour can be deformed at will across the real axis, and the integral can be calculated by the saddle-point method. We start by extending the integral

$$u_2 \equiv -\frac{1}{2}\int d\nu\, e^{i\nu(\varphi-\pi/2)}\,\frac{H_\nu^{(2)}(q\rho)H_\nu^{(1)}(qa) - H_\nu^{(2)}(qa)H_\nu^{(1)}(q\rho)}{H_\nu^{(1)}(qa)} \tag{3.4.21}$$

along a contour $D'$ that passes around the curve $\operatorname{Re}\alpha = 0$ so as to stay to the right of $\operatorname{Re}(\alpha + i\pi\nu) = 0$ ($\alpha = \sqrt{\nu^2 - (qa)^2} - \nu \ln(\nu + \sqrt{\nu^2 - (qa)^2})/qa$), and which then follows a curve containing the zeroes of $H_\nu^{(1)}(q\rho)$:

$$\operatorname{Re}\left[\sqrt{\nu^2 - (q\rho)^2} - \nu \ln \nu + \frac{\sqrt{\nu^2 - (q\rho)^2}}{q\rho}\right] = 0. \tag{3.4.22}$$

According to (3.4.7), in this region $H_\nu^{(2)}(q\rho)$ decreases exponentially, and the first term in the numerator of (3.4.21) contributes nothing, because it is an entire function in $\nu$. As for the second term, on the left-side portion of $D'$ the factor $H_\nu^{(2)}(qa)$ is exponentially decreasing, whereas on the right-side portion the variable can follow a path through the zeroes of $H_\nu^{(1)}(q\rho)$, which are on (3.4.22). On this portion of the contour the factor $\exp(i\nu(\varphi - \pi/2))$ takes care

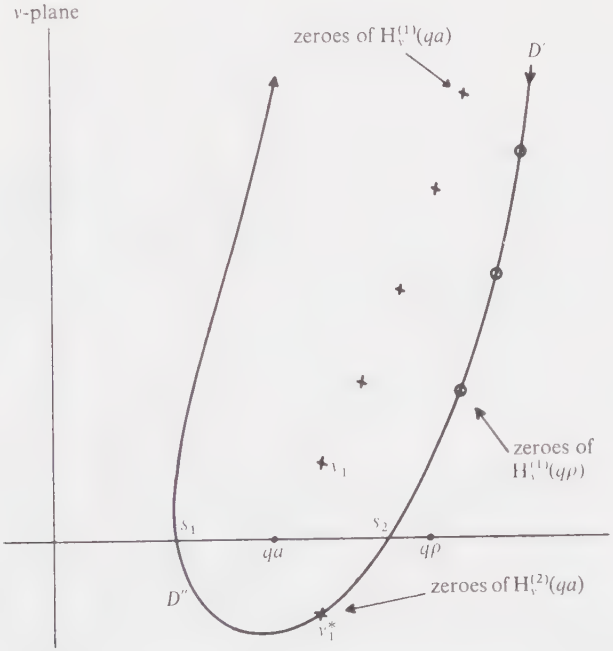



Figure 51 Path of integration for calculating geometric optics

of the exponential decrease, and this part of the contour can be connected to a path $D''$ that crosses the real axis at a saddle point $s_1 < qa$, proceeds to the first zero $\nu^*$ of $H_\nu^{(2)}(qa)$, and then climbs over another saddle point $s_2 > qa$ to reach the valley containing the zeroes of $H_\nu^{(1)}(q\rho)$. We next determine $s_i$ in the case $|s_i - qa| > (qa)^{1/3}$; the other case will be considered in (3.4.34). The integrand of $u_2$ on the left-side portion of $D''$ looks like

$$\frac{1}{\sqrt{2\pi}} e^{i\nu(\varphi-\pi/2)}((q\rho)^2 - \nu^2)^{-1/4} e^{i\pi/4}$$

$$\cdot \exp\left\{ i\left[ -2(q^2a^2 - \nu^2)^{1/2} + (q^2\rho^2 - \nu^2)^{1/2} \right.\right.$$

$$\left.\left. + 2\nu \arccos\frac{\nu}{aq} - \nu \arccos\frac{\nu}{q\rho} \right]\right\}, \tag{3.4.23}$$

because of the asymptotic properties of the Hankel functions. The $\nu$ dependence is dominated by the exponent, and the saddle point $s_1$ is at the position where the derivative by $\nu$ vanishes,

$$\varphi - \frac{\pi}{2} + 2 \arccos\frac{s_1}{qa} - \arccos\frac{s_1}{q\rho} = 0. \tag{3.4.24}$$

With the notation $s_1 \equiv qp$, the integrand becomes

$$\frac{\exp(-iq\sqrt{a^2 - p^2})}{\sqrt{q}(\rho^2 - p^2)^{1/4}} \exp\left\{ iq\left[ \sqrt{\rho^2 - p^2} - \sqrt{a^2 - p^2} \right]\right\} \frac{e^{i\pi/4}}{\sqrt{2\pi}} \tag{3.4.25}$$

at the saddle point. We can now recognize $p$ as the impact parameter of the ray that is reflected to the observation point according to the laws of geometric optics, and that the phase of (3.4.25) is exactly that of the optical path having this reflection (Figure 52). Therefore this is the contribution from reflected light.

On the right side portion of $D''$, $H_\nu^{(2)}(qa)\, H_\nu^{(1)}(qa) \sim -1$, and the asymptotic form of the integrand simplifies to

$$-\frac{e^{-i\pi/4}}{\sqrt{2\pi}} \frac{\exp\left\{ i\left( \nu\left( \varphi - \frac{\pi}{2} - \arccos\frac{\nu}{q\rho} \right) + \sqrt{\rho^2 q^2 - \nu^2} \right)\right\}}{\sqrt[4]{(q\rho)^2 - \nu^2}}. \tag{3.4.26}$$

The saddle point $s_2$ is at

$$\varphi - \frac{\pi}{2} - \arccos\frac{s_2}{q\rho} = 0, \tag{3.4.27}$$

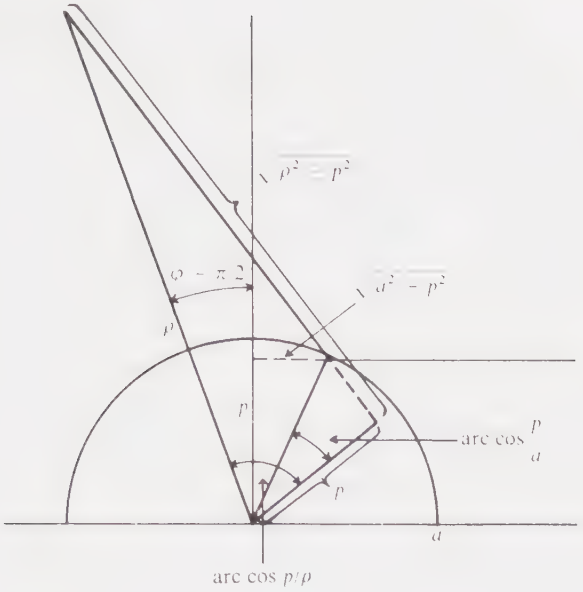



Figure 52 The reflected light according to geometric optics

and, with $p = \nu/q$, the integrand has the value

$$-\frac{e^{-i\pi/4}}{\sqrt{2\pi}} \frac{\exp(iq\sqrt{\rho^2 - p^2})}{\sqrt[4]{(q\rho)^2 - \nu^2}} \tag{3.4.28}$$

at the saddle point. Consequently $p$ is the impact parameter of the incident ray, and the phase of (2.4.28) is simply $\exp(-iqx)$ (Figure 53).

To calculate the incoming wave precisely, it is necessary to go through the usual manipulations of the method of steepest descent:

$$\int d\nu\, A(\nu) e^{i\alpha(\nu)} \cong A(s_2) e^{i\alpha(s_2)} \int_{-\infty}^{\infty} d\nu \exp\left\{ i\alpha''(s_2) \frac{(\nu - s_2)^2}{2} \right\}$$

$$= A(s_2) \sqrt{\frac{-2\pi}{i\alpha''(s_2)}}\, e^{i\alpha(s_2)}; \tag{3.4.29}$$

since

$$\begin{aligned} A(s_2) &= \frac{-e^{-i\pi/4}}{\sqrt{2\pi}((q\rho)^2 - s_2)^{1/4}}, \\ \alpha''(s_2) &= ((q\rho)^2 - s_2)^{-1/2}, \end{aligned} \tag{3.4.30}$$

we find exactly $\exp(iq\sqrt{\rho^2 - p^2}) = \exp(-iqx)$. The corrections to the asymptotic expression can be calculated systematically, and are $O(\rho^{-2})$.

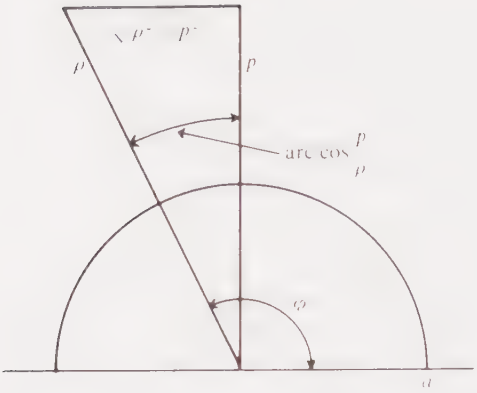



Figure 53 The direct radiation according to geometric optics

Similarly, the contribution $u_{2r}$ of the reflected light can be calculated accurately, to reveal that there is an amplitude factor that controls the spreading out of the reflected light. In (3.4.30) the only changes in $A$ are signs, but, because of (3.4.24),

$$\alpha''(v) = 2(a^2q^2 - v^2)^{-1/2} - (\rho^2q^2 - v^2)^{-1/2}. \tag{3.4.31}$$

Together with $u_2 \cong \exp(-iqx) + u_{2r}$, this makes

$$u_{2r} = \left(\frac{a^2 - p^2}{\rho^2 - p^2}\right)^{1/4}\left(2 - \left(\frac{a^2 - p^2}{\rho^2 - p^2}\right)^{1/2}\right)^{-1/2}$$
$$\cdot \exp\{iq[\sqrt{\rho^2 - p^2} - 2\sqrt{a^2 - p^2}]\}. \tag{3.4.32}$$

If $\rho \gg a$, then $\sqrt{a^2 - p^2} = a\cos\varphi/2 \equiv a\sin\vartheta/2$, where $\vartheta = \pi - \varphi$ is the scattering angle. Then

$$u_{2r} \cong \sqrt{\frac{a\sin\vartheta/2}{2\rho}}\exp\left\{i\left(q\rho - 2qa\sin\frac{\vartheta}{2}\right)\right\}. \tag{3.4.33}$$

The angular distribution $\rho|u_{2r}|^2\, d\vartheta = dp$ is exactly what the geometric law of reflection predicts. Although our derivation works only if $\varphi > \pi/2$, the same result can be obtained for $\varphi < \pi/2$ by making minor changes, and by symmetry $u(\varphi) = u(-\varphi)$.

**The Boundary of the Shadow** (3.4.34)

The validity of the result of the previous paragraph relied on

(a) $\varphi > \pi/2$, so that the exponential factor decreases in the upper half-plane;
(b) staying away from the region where $\varphi \to \pi$ and $\rho \to \infty$, in which the creeping waves are insufficiently attenuated; and
(c) $|p - a| > a(qa)^{-2/3}$, so that the saddle point is sufficiently far from $qa$.

At the boundary of the shadow, assumption (c) fails, and we would have to assume that $\rho$ is not too large, in order to preserve (b). Granting that, Equation (3.4.21) could still be used, although both saddle points now move toward $qa$. The integration contour is divided into $D_1$, reaching from $aq + i\infty$ to $aq$; $D_2$, from $aq$ to $q\rho_1 < q\rho$; and $D_3$, from $q\rho_1$ to infinity along (3.4.22). The integrand decreases exponentially fast along $D_1$ and $D_3(\varphi > \pi/2)$, and on $D_2$ the ratio $H^{(2)}/H^{(1)} = -1 + 2J/H^{(1)}$ can be replaced with $-1$. The other contributions are straightforward to estimate as being $O((qa)^{1/3}(\varphi - \varphi_0))$, where $\varphi_0$ is the angle such that $(\rho, \varphi_0)$ is on the geometric boundary of the shadow, and we are interested in the values $|\varphi - \varphi_0| \sim (qa)^{-1/2}$.

Under these circumstances we are thus led back to

$$u_2 = \frac{e^{-i\pi/4}}{\sqrt{2\pi}} \int_{qa}^{q\rho_1} dv \frac{\exp\left\{iv\left(\varphi - \frac{\pi}{2} - \arccos\frac{v}{q\rho}\right) + i\sqrt{\rho^2q^2 - v^2}\right\}}{[(q\rho)^2 - v^2]^{1/4}} \tag{3.4.35}$$

(cf. (3.4.26)), since the dominant part of the integral ought to come from the vicinity of $qa$, and we have used the proper asymptotic form of $H_\nu^{(1)}(q\rho)$ for that region. By the substitution

$$v \equiv q\rho \sin\omega, \qquad \varphi_0 > \omega > \varphi_1 > \frac{\pi}{2} \tag{3.4.36}$$

($a = \rho \sin\varphi_0$, $\rho_1 = \rho \sin\varphi_1$), (3.4.35) becomes

$$u_2 = \frac{e^{-i\pi/4}}{\sqrt{2\pi}} \int_{\varphi_1}^{\varphi_0} d\omega(|q\rho\cos\omega|)^{1/2} e^{iq\rho[(\varphi-\omega)\sin\omega - \cos\omega]}. \tag{3.4.37}$$

The function in the brackets [ ] in the exponent has its minimum at $\omega = \varphi$, and its leading term as $q\rho \to \infty$ can be approximated by a parabolic function,

$$[\ \ ] \cong -\cos\varphi\left(1 + \frac{(\omega - \varphi)^2}{2}\right).$$

If we write

$$\rho\cos\varphi = x, \qquad -\frac{iqx(\varphi - \omega)^2}{2} = \frac{i\pi\tau^2}{2}, \tag{3.4.38}$$

then we discover that

$$u_2 = e^{-iqx}\frac{e^{-i\pi/4}}{\sqrt{2|x|q}} \int_{\sqrt{q|x|/\pi}(\varphi-\varphi_0)}^{\sqrt{q|x|/\pi}(\varphi-\varphi_1)} d\tau\, e^{i\pi\tau^2/2}$$

$$\cdot\left[q|x|\cos\tau\sqrt{\frac{\pi}{q|x|}} + q\rho\sin\varphi\sin\tau\sqrt{\frac{\pi}{q|x|}}\right]^{1/2}. \tag{3.4.39}$$

As $q|x| \to \infty$, the upper limit of integration goes to infinity and $[\ ]^{1/2}$ approaches $\sqrt{q|x|}$; but unlike what we had earlier, the lower limit of integration remains finite if we are in the region $|\varphi - \varphi_0| \sim (q|x|)^{-1/2}$. With Fresnel's integral

$$F(x) = \int_{-\infty}^{x} d\tau\, e^{i\pi\tau^2/2} \tag{3.4.40}$$

we can simply write

$$u_2 = e^{-iqx}\frac{e^{-i\pi/4}}{\sqrt{2}}\left(F(\infty) - F\left((\varphi - \varphi_0)\sqrt{\frac{q|x|}{\pi}}\right)\right). \tag{3.4.41}$$

In other words, $F$ specifies how the incident wave $\exp(-iqx)$ dies down as it passes from light ($\varphi - \varphi_0 < 0$, $F(\infty) = \sqrt{2}\exp(i\pi/4)$) to shadow ($\varphi - \varphi_0 > 0$, $u_2 \to 0$). At the geometric edge of the shadow ($\varphi = \varphi_0$) it has only half the original amplitude, because $F(0) = F(\infty)/2$, and as it passes from that point into the light, it oscillates according to Cornu's spiral around the original amplitude (cf. (3.3.10)).

**The Frauenhofer Region** (3.4.42)

Finally, we undertake a study of the region $\rho > (qa)a$, where the shadow has already faded out. If $|\varphi - \pi| > (qa)^{-1/3}$, then we know from (3.4.19) that the wave is simply composed of the incident wave plus a wave reflected according to geometric optics. If $|\varphi - \pi| < (qa)^{-1/3}$, then the facts discussed in (3.4.19) do not suffice to cause attenuation of the creeping waves, and we are forced to take drastic steps. We shall rely on the uniformly convergent expansion in the upper $\nu$-plane,

$$\frac{e^{-i\nu\pi/2}}{\sin \nu\pi} = -2ie^{i\nu\pi/2}\sum_{m=0}^{\infty} e^{2i\pi m\nu}. \tag{3.4.43}$$

Note that the condition for the contributions from the residues to decrease sufficiently fast is now

$$|\varphi - \pi| + \arccos\frac{a}{\rho} < \pi\left(2m + \frac{1}{2}\right) - (aq)^{-1/3} \tag{3.4.44}$$

(cf. (3.4.17)), which is always true for $m \geq 1$, so we only need to worry about the contribution $u_0$ with $m = 0$ in the sum. To take care of that, we again make the partition $D' = D_1 \cup D_2 \cup D_3$ of the earlier paragraph, and see that the main contribution again comes from $D_2$ with $H^{(2)}/H^{(1)} \to -1$. We shall only calculate the main contribution, as an estimate of the others shows that they become negligible in comparison with it as $\rho \to \infty$. Choosing $\rho_1 = \rho/2$,

we can write the $m = 0$ term of the integral over $D_2$ of the asymptotic form of the integrand (3.4.6) as

$$u_{20} = -\frac{e^{-i\pi/4}}{\sqrt{2\pi}} \int_{qa}^{q\rho/2} \frac{dv}{((q\rho)^2 - v^2)^{1/4}}$$

$$\cdot \exp\left\{iv\left(\vartheta + \frac{\pi}{2} - \arccos\frac{v}{q\rho}\right) + i\sqrt{(q\rho)^2 - v^2}\right\} + (\vartheta \to -\vartheta)$$

$$= -e^{i(q\rho - \pi/4)}\sqrt{\frac{q\rho}{2\pi}} \int_{a/\rho}^{1/2} d\tau (1 - \tau^2)^{-1/4} e^{iq\rho\alpha(\tau)} + (\vartheta \to -\vartheta),$$

$$\alpha(\tau) = \vartheta\tau + \tau \arcsin \tau + \sqrt{1 - \tau^2} - 1. \tag{3.4.45}$$

We have introduced the scattering angle $\vartheta = \pi - \varphi > 0$ and substituted from $\tau \equiv v/q\rho$. Since the derivative of the exponent $\alpha'(\tau) = \vartheta + \arcsin \tau$ never vanishes, we may make the usual integration by parts,

$$\int_{a/\rho}^{1/2} d\tau\, A(\tau) e^{iq\rho\alpha(\tau)} = \int_{a/\rho}^{1/2} d\tau \frac{A(\tau)}{iq\rho\alpha'(\tau)} \frac{\partial}{\partial\tau} e^{iq\rho\alpha(\tau)}$$

$$= \frac{A(\tau)e^{iq\rho\alpha(\tau)}}{iq\rho\alpha'(\tau)}\Bigg|_{a/\rho}^{1/2} + O((q\rho)^{-2}) \tag{3.4.46}$$

to determine the asymptotic properties. For values $\vartheta \sim (qa)^{-1}$ but $\vartheta \gg a/\rho$, i.e., $\rho \gg aaq$, we thereby find that

$$u_{20} \approx \frac{e^{iq\rho}}{\sqrt{\rho}} e^{-i\pi/4} \sqrt{\frac{2}{\pi q}} \frac{\sin qa\vartheta}{\vartheta}. \tag{3.4.47}$$

This is the same as the diffraction pattern at a slit or cylinder as calculated with the older theory, in which Formula (1.2.36) is used with $u$ in the lighted regions replaced with the incident wave. However, the older procedure gives no indication of the errors incurred, whereas the above theory is able to specify corrections to the asymptotic formulas.

If $\rho \gg aqa$, then $u$ is of the form

$$u \cong e^{-iqx} + \frac{e^{iq\rho}}{\sqrt{\rho}} f(\vartheta) \equiv u_{\text{in}} + u_{\text{scat}}. \tag{3.4.48}$$

If the wave is normally incident ($k = 0$ in (3.2.19; 3) and $q = \omega$), $f$ determines the

**Scattering Cross-Section** (3.4.49)

$$\sigma(\vartheta) \equiv \lim_{\rho\to\infty} \rho \frac{(T^{0\rho})_{\text{scat}}}{(T^{00})_{\text{in}}} = |f(\vartheta)|^2,$$

where the energy-momentum tensor is to be calculated respectively with $u_{\text{scat}}$ or $u_{\text{in}}$.

**Remarks** (3.4.50)

1. The cross-section $\sigma$ is the radiated energy per unit angle divided by the incident energy per unit length. It has the dimensions of a length. We are using the energy current from (3.1.24) for the scalar field $u$ (see Problem 1); the electromagnetic problem has a somewhat more complicated formulation.
2. In the limit $qa \gg 1$ and $\rho \to \infty$, we find that

$$|f(\vartheta)|^2 = \frac{a}{2}\sin\frac{\vartheta}{2} \qquad \text{if } \vartheta > (qa)^{-1/3},$$

$$|f(\vartheta)|^2 = \frac{2}{\pi q}\left|\frac{\sin qa\,\vartheta}{\vartheta}\right| \text{ if } \vartheta \sim (qa)^{-1};$$

in words, in addition to the geometrically scattered light there is an extra forward maximum, originating in the waves that permeate the shadow. Interestingly enough, both components have the geometric cross-section $2a$; but there is hardly any contribution from $(qa)^{-1} \ll q < (qa)^{-1/3}$:

$$2\int_{(qa)^{-1/3}}^{\pi} d\vartheta \frac{a}{2}\sin\frac{\vartheta}{2} = 2a(1 + O((qa)^{-2/3}))$$

$$2\int_{0}^{(qa)^{-1/3}} d\vartheta \frac{2}{\pi q}\left|\frac{\sin qa\vartheta}{\vartheta}\right|^2 = 2a(1 + O((qa)^{-1/3})).$$

Hence the total cross-section is twice the geometric cross-section. The significance is that not only is the directly incident light reflected, but also the light passing by at a distance $a$ is somewhat deflected.

3. In the limit $qa \ll 1$, the scattering cross-section is strongly dependent on the polarization. Until now we have always assumed that $u_{|\partial N} = 0$, for which (3.4.4; 2) makes the scattering cross-section

$$\sigma(\vartheta) = \frac{\pi a}{2qa|\ln qa|^2}.$$

This can exceed the geometric cross-section by an arbitrary amount, even though it obviously vanishes if $q$ is fixed and $a \to 0$.

4. If $du_{|\partial N} = 0$, then the Fourier coefficients in (3.4.2) are replaced with $J_n'(qa)/H_n^{(1)\prime}(qa)$, and if $qa \to 0$, then the terms with $n = 0, \pm 1$ contribute:

$$\sigma = \frac{\pi(qa)^3}{8}a(1 - 2\cos\vartheta)^2.$$

The total cross-section $a(3\pi^2/4)(qa)^3$ is much smaller than the geometric cross-section.

5. In electromagnetism Remark 3 applies to a field **E** parallel to the $z$-axis, and Remark 4 to **E** perpendicular to the $z$-axis. The physical reason for the greater cross-section of Remark 3 is that it is easier to move charges along the axis of the cylinder than transversely.

The example we have discussed has shown how complicated diffraction phenomena can be. Even Cartan's formalism, which makes the algebraic complications trivial, is not very effective for these analytical problems. However, the integral representation (3.4.6) is versatile enough that the whole range of diffraction phenomena can be derived from it, like rabbits pulled from a magician's hat. In general interference of waves renders the solution chaotic. Only in some limits the simple laws of geometrical optics emerge.

**Problems** (3.4.51)

1. Derive (3.4.49) for the energy-momentum forms (3.1.24) with

$$J_\alpha = u_{,\alpha}, \qquad u_{\text{in}} = e^{-iq(x+t)}, \qquad u_{\text{scat}} \cong \frac{e^{iq(\rho-t)}}{\sqrt{\rho}} f(\varphi).$$

2. Derive the optical theorem

$$\sigma_{\text{tot}} = \int_{-\pi}^{\pi} d\vartheta\, \sigma(\vartheta) = -2\sqrt{\frac{2\pi}{q}}\, Re(e^{i\pi/4} f(0)),$$

where $f$, as in (3.4.48) is defined by the asymptotic form of

$$u = \sum_{n=-\infty}^{\infty} e^{in((\pi/2)-\vartheta)} \left[ \mathrm{J}_n(q\rho) - \frac{\mathrm{J}_n(qa)}{\mathrm{H}_n^{(1)}(qa)} \mathrm{H}_n^{(1)}(q\rho) \right] \overset{\rho\to\infty}{\longrightarrow} e^{-iqx} + f(\vartheta) \frac{e^{iq\rho}}{\sqrt{\rho}}.$$

Use the formula

$$\frac{\mathrm{J}_n}{\mathrm{H}_n^{(1)}} = \frac{1}{2} \frac{\mathrm{H}_n^{(1)} + \mathrm{H}_n^{(2)}}{\mathrm{H}_n^{(1)}} \equiv \frac{1}{2}(1 - e^{2i\delta_n}), \qquad \delta_n \text{ real.}$$

3. Calculate the scattering cross-sections given in (3.4.50; 3) and (3.4.50; 4).

**Solutions** (3.4.52)

1. $$\frac{(T^{0\rho})}{(T^{00})} = \frac{\dot{u}u_{,\rho}}{\frac{1}{2}(\dot{u}^2 + |\nabla u|^2)}, \quad \frac{(T^{0\rho})_{\text{scat}}}{(T^{00})_{\text{in}}} = \rho \frac{q^2 |f(\varphi)|^2/\rho}{q^2} = |f(\varphi)|^2.$$

Since $|u(t)|^2 = (\mathrm{Re}\, u(t))^2 + (\mathrm{Im}\, u(t))^2 = (\mathrm{Re}\, u(t))^2 + (\mathrm{Re}\, u(t + \pi/2\omega))^2$, the use of the absolute value is equivalent to a time-average.

2. $$\frac{f(\vartheta)}{\sqrt{\rho}} e^{iq\rho} = \frac{1}{2} \sum_{n=-\infty}^{\infty} e^{in((\pi/2)-\vartheta)} (e^{2i\delta_n} - 1) \mathrm{H}_n^{(1)}(q\rho).$$

From the asymptotic expression

$$H_n^{(1)}(q\rho) \cong \sqrt{\frac{2}{\pi q\rho}}\, e^{i(q\rho-(n\pi/2)-(\pi/4))}$$

one finds that

$$f(\vartheta) = \frac{e^{-i\pi/4}}{\sqrt{2\pi q}} \sum_{n=-\infty}^{\infty} e^{-in\vartheta}(e^{2i\delta_n} - 1), \qquad \sigma = \int_{-\pi}^{\pi} |f(\vartheta)|\,d\vartheta = \frac{4}{q} \sum_{n=-\infty}^{\infty} \sin^2 \delta_n,$$

$$\mathrm{Re}(e^{i\pi/4} f(0)) = \frac{-1}{\sqrt{2\pi q}} \sum_{n=-\infty}^{\infty} (1 - \cos^2 \delta_n + \sin^2 \delta_n).$$

3. The cross-section of Remark 3 follows from (3.4.4; 2). As for Remark 4,

$$u = \sum_{n=-\infty}^{\infty} e^{in((\pi/2)-\vartheta)} \left( J_n(q\rho) - \frac{J_n'(qa)}{H_n^{(1)\prime}(qa)} H_n^{(1)}(q\rho) \right),$$

$$f(\vartheta) = \sqrt{\frac{2}{\pi q}}\, e^{-i\pi/4} \left( \frac{J_0'(qa)}{H_0^{(1)\prime}(qa)} + 2\cos\vartheta \frac{J_1'(qa)}{H_1^{(1)\prime}(qa)} + \cdots \right)$$

$$\cong \sqrt{\frac{2}{\pi q}}\, e^{-i\pi/4} \left( \frac{qa}{2} \right)^2 (1 - 2\cos\vartheta) i\pi.$$

# Gravitation 4

## 4.1 Covariant Differentiation and the Curvature of Space

*The covariant derivative defines the rate of change of a tensor field in the direction of a vector. Covariant derivatives in two different directions do not in general commute; their commutator determines the curvature of space.*

In field theory one has to deal with derivatives of vector fields and in modern theories there appear quantities which are vectors not in space-time but in an internal space. In both cases one deals with vector bundles where vectors at different points are not canonically oriented towards each other. A chart independent notion of a derivative requires an additional structure, the so-called connection. It will be the subject of this chapter. As one hopes that eventually space, time and internal space will turn out to be only different directions in a unifying entity we start with some definitions which allow us to treat both cases in the same way.

**Definition** (4.1.1)

A **section** of a vector bundle $V$ with basis $B$ and projection $\Pi$ is a map $\Phi$: $B \to V$ with $\Pi \circ \Phi = 1_B$. The set of sections is denoted by $S_0(V)$.

**Examples** (4.1.2)

1. The charged scalar field (3.1.21; 2) $\Phi$ has two components $\varphi^1$, $\varphi^2$ and can be considered as a section of the bundle $B \times \mathbb{R}^2$ where $B$ is space-time.

It associates to $x \in B$ the point $(x; b_1\varphi^1(x) + b_2\varphi^2(x)) \in V$ if $b_{1,2}$ are a basis of $\mathbb{R}^2$ which is the fibre in this case.

2. Besides the charged pions $\pi^{\pm}$ there exists the neutral pion $\pi^0$ with about the same mass. Their description requires a 3-component pseudo-scalar field $\varphi^i$, $i = 1, 2, 3$ where $\varphi^1 \pm i\varphi^2$ describe $\pi^{\pm}$ and $\varphi^3$ describes $\pi^0$. Thus the pion field is a section in a vector bundle with fibre $\mathbb{R}^3$, the so-called isospin space. There are many examples of similar internal degrees of freedom in particle physics.
3. Vector fields (resp. 1-forms) are sections in the bundles $T(B)$ (resp. $T^*(B)$), $S_0(T^*(B)) = \mathscr{T}^0_1(B)$, etc. In our case $B$ will be space-time, dim $B = 4$. Thus the fibres in $T(B)$ will always be $\mathbb{R}^4$. Similarly tensor fields are sections of the tensor bundles.

We shall encounter only situations where the fibres have in addition to their structure of a vector space another important property which deserves a name.

**Definition** (4.1.3)

A nondegenerate bilinear map (or sequilinear form for complex bundles) $S_0(V) \times S_0(V) \to C(B)$: $(\Phi, \Psi) \to \langle \Phi | \Psi \rangle$ is called **fibre metric**.

**Remark** (4.1.4)

$S_0(V)$ is a module over $C(B)$ and bilinearity implies $\langle f\Phi | g\Psi \rangle = fg\langle \Phi | \Psi \rangle \; \forall f, g \in C(B)$. By nondegenerate we mean $\langle \Phi | \Psi \rangle = 0 \; \forall \Phi \in S_0 \Rightarrow \Psi = 0$ at every point of $B$.

**Examples** (4.1.5)

1. In the Lagrangian for the charged scalar field or the pion field appears the combination $\sum_i \varphi^{i2}(x)$. It corresponds to the usual metric in $\mathbb{R}^n$ and will be taken as fibre metric for this bundle.
2. On a Riemannian space $M$ the metric defines a fibre metric in the bundle $T(M)$. We have seen how it can be extended to $T^*(M)$ and the tensor bundles.
3. On any manifold $M$ the canonical 2-form on $T^*(M)$ defines a fibre metric in the bundle $T(T^*(M))$. Note that the fibre metric need not be positive.

Every element of a vector space can be expanded in terms of a basis $\{b_i\}$, but for a vector bundle a basis for the sections may not exist globally but only locally. In any case the metric is locally determined by its action on a basis and for a symmetric metric it is expedient to use a basis where the metric assumes the normal form

$$\langle b_i | b_j \rangle = \eta_{ij} \qquad \text{with} \qquad \eta_{ij} = \begin{cases} \pm 1 & \text{for } i = j, \\ 0 & \text{for } i \neq j, \end{cases}$$

("orthonormal" or simply "orthogonal basis"). This requirement does not fix the $b_i$ uniquely, $b_i = b_k L^k{}_i$ is also an orthogonal basis if $L^t \eta L = \eta$. Even though the $\eta_{ik}$ do not depend on $x \in B$ the $L$ may do so.

**Definition** (4.1.6)

Let $\{b_i\}$ be a local basis of sections in a vector bundle with fibre metric $\langle b_i | b_j \rangle = \eta_{ij}$. The transformations $b_i \to b_k L^k{}_i$ with $L^t \eta L = \eta$ are called **gauge transformations**. The $L^k{}_i(x)$ form the same group for all $x \in B$, the so-called **gauge group** $G$ of the bundle. If the bundle is trivializable and $L: B \to G$ is a constant map one speaks of global gauge transformations.

**Examples** (4.1.7)

1. For the fields $\Phi = \sum b_i \varphi^i$ with fibre metric $\langle \Phi | \Phi \rangle = \sum_{i=1}^{n} \varphi^{i^2}$ the gauge group $G$ is $O(n)$.
2. In a pseudo-Riemannian space with signature $(n, m)$ ($\eta$ has $n$ positive and $m$ negative eigenvalues) the gauge group is $O(n, m)$.
3. In $T(T^*(M))$ with the canonical 2-form as fibre metric the gauge group is the symplectic group.

**Remark** (4.1.8)

There are groups which cannot be characterized by the invariance of a quadratic form. To treat them as a gauge group one has to ascend to the next level of abstraction and define principal fibre bundles where the fibres are the group manifold itself. In our cases the gauge group will always come from the invariance of a fibre metric and we shall not need this construction.

For an exterior derivative of a section we still need a notion which combines sections and $p$-forms.

**Definition** (4.1.9)

The sections over the vector bundle $B \times V \otimes_\pi \wedge_p T^*(B)$ are called **$p$-form valued sections** (or vector valued $p$-forms), their set is denoted by $S_p$.† They are a linear space and the exterior product $\wedge$ maps $E_p \times S_q$ into $S_{p+q}$. Thus $\Phi \in S_p$ can be written $\Phi = \sum_i b_i v^i$ with $v^i \in E_p$ and $b_i$ the basis in $S_0$. Given a vector field $u \in \mathcal{T}^1_0(B)$ we define an inner product $i_u: S_p \to S_{p-1}$ by

$$i_u \Phi = \sum_j b_j i_u v^j.$$

† $\otimes_\Pi$ denotes a product of bundles with the same basis. In it the fibres are the tensor product of individual fibres and the basis is the common basis. $\wedge_p T^*(B)$ is the $p$-fold antisymmetric product of the $T^*(B)$.

So far there is no chart-independent notion of parallelism between fibres at different points. To define the derivative of a section $\Phi$ we need a prescription how to parallel transport $\Phi(x)$ to $x + dx$ such that $\Phi(x + dx) - \Phi(x)$ can be defined. This prescription will be expressed on the infinitesimal level by the connection. We shall wonder later whether and how this parallelism can be extended to the local or global levels. First we list some formal desiderata for the derivative and then successively narrow it down by requiring that it conserve further structures. To start with we want a derivative $D_u$ of a section in direction of a vector field $u$ to be linear in $u$ so that one ought to be able to write it $i_u D$, $D\colon S_0 \to S_1$, hence we concentrate on $D$.

**Definition** (4.1.10)

A **covariant exterior derivative** $D$ is a map $S_p \to S_{p+1}$ with the properties:

(i) $D(\Phi_1 + \Phi_2) = D\Phi_1 + D\Phi_2, \qquad \Phi_i \in S_p;$
(ii) $D(\Phi \wedge v) = (D\Phi) \wedge v + (-)^p \Phi \wedge dv, \qquad \Phi \in S_p, \qquad v \in E_q.$

The 1-forms $\omega^k{}_i$ appearing in the expansion of the covariant exterior derivative of a local basis $b_i \in S_0$

$$Db_i = b_k \omega^k{}_i, \qquad \omega^k{}_i \in E_1,$$

are called the **linear connection or gauge potential**.

**Remarks** (4.1.11)

1. Through the rules (i) and (ii) $D$ is completely determined by the $\omega$'s. For instance, with the natural basis $dx^\alpha$ in $E_1$ we can write

$$\Phi = \frac{1}{p!} \sum_{i,(\alpha)} b_i \varphi^{i(\alpha)}\, dx^{\alpha_1} \wedge \cdots \wedge dx^{\alpha_p}$$

and $D\Phi$ becomes with the rules (4.1.10)

$$D\Phi = \frac{1}{p!} \sum_{i,(\alpha)} b_i (d\varphi^{i(\alpha)} + \omega^i{}_k \varphi^{k(\alpha)}) \wedge dx^{\alpha_1} \wedge \cdots \wedge dx^{\alpha_p}.$$

2. Since the $b_i$ need not be defined globally the $\omega$ will not be either, even though $D$ exists globally.
3. $\omega$ can be considered as element of $S_1(L(F))$, a 1-form valued section of the bundle over the domain of the $b_i$ with the linear transformations of the fibres $F$ as fibres. In this sense one might use the shorthand $D = d + \omega\wedge$. However, one has to keep in mind that under a change of the basis in $F$ the $\omega$'s do not transform the way elements of $S_1(L(F))$ do.

**Transformation of the Gauge Potential under Gauge Transformations** (4.1.12)

Under a change of the local basis $b \to bL$, $L \in S_0(L(F))$ the $\omega$ transform as

$$\omega \to L^{-1}\omega L + L^{-1}\,dL.$$

**Remarks** (4.1.13)

1. $dL \in S_1(L(F))$ is to be understood as the matrix with elements $dL^i{}_k$ if the change of basis is $b_k \to \bar{b}_k = b_i L^i{}_k$. Thus, with index notation

$$\bar{\omega}^i{}_k = (L^{-1})^i{}_j\,\omega^j{}_l L^l{}_k + (L^{-1})^i{}_j\,dL^j{}_k.$$

2. The transformation law follows from the rules (4.1.10)

$$\bar{b} = bL \Rightarrow D\bar{b} = (Db)L + bdL = b\omega L + b\,dL = \bar{b}(L^{-1}\omega L + L^{-1}\,dL).$$

3. Looked at from the point of view of the components $\varphi^i$ of $\Phi = \sum b_i \varphi^i \in S_0$ one can say that the components $d\varphi^i + \omega^i{}_k \varphi^k$ of $D\Phi$ transform like the $\varphi^i$ under the gauge transformation $\varphi^i \to \bar{\varphi}^i = (L^{-1})^i{}_k \varphi^k$. The inhomogeneous term $(dL^{-1})\Phi$ is cancelled by the corresponding term in the transformed $\omega$ since $L^{-1}L = 1 \Rightarrow dL^{-1}L = -L^{-1}\,dL$.

**Examples** (4.1.14)

1. It may happen that one basis $b_i$ is covariant constant, $\omega^i{}_k = 0$. In another basis $\bar{b}_k = b_i L^i{}_k$ we have then

$$D\bar{b}_k = \bar{b}_j (L^{-1})^j{}_i\,dL^i{}_k.$$

   If there is no global basis it is therefore not clear whether there is a $D$ with $\omega = 0$ on all of $B$.

2. The model (3.1.21; 2) of a charged field can be written in terms of two real fields $\varphi^1 = \sqrt{\rho}\cos S$, $\varphi^2 = \sqrt{\rho}\sin S$ or one complex field $\Phi = \varphi^1 + i\varphi^2 = \sqrt{\rho}\,e^{iS}$. To get a vector bundle we have to allow $\rho$ to vary over $\mathbb{R}^+$. In this case we have the fibres $F$ equal to $\mathbb{R}^2$ or $C$ and the bundle $V = \mathbb{R}^4 \times F$. In the Lagrangian (3.1.20) occurred the combination $d\varphi + iA\varphi$ (or $d\varphi^i + \omega^i{}_k \varphi^k$, $\omega^i{}_k = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right) A(x)$) which is a covariant exterior derivative in this bundle. $S \to S + \Lambda$ corresponds to a change of basis in which case $A$ undergoes the gauge transformation $A \to A + d\Lambda$. Here the gauge transformations and the connection involve only the matrix $\left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$ and therefore the $L$ cancel out in $L^{-1}\omega L$.

So far our motivation for $D$ was more formal. Its geometrical significance becomes more apparent when one defines the covariant derivative in the direction of a vector field over $B$.

**Definition** (4.1.15)

Given a vector field $u \in \mathcal{T}^1_0(B)$ the **covariant Lie derivative** is defined by

$$\mathcal{L}_u = i_u \circ D + D \circ i_u.$$

It maps $S_p$ into $S_p$. The particular case $p = 0$ is called **covariant derivative** and is denoted by $D_u \equiv i_u \circ D$.

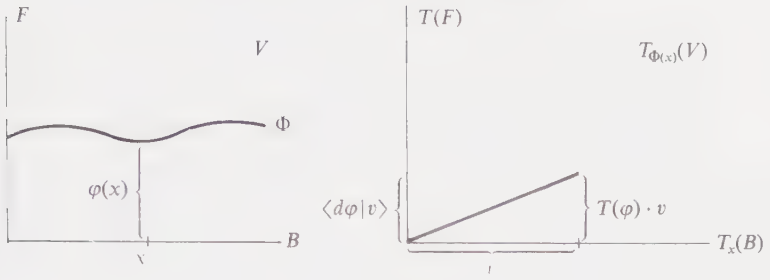
**Remarks** (4.1.16)

1. Remember how the inner product $i_u$ was defined. We have more explicitly for $\varphi = \sum_i b_i \varphi^i \in S_0$ the expression
$$\mathcal{L}_u \Phi = D_u \Phi = \sum_i b_i((d\varphi^i|u) + (\omega^i{}_k|u)\varphi^k).$$
2. For $\mathcal{L}_u$ like for $L_u$ the Leibniz rule, $\mathcal{L}_u(v \wedge \Phi) = (\mathcal{L}_u v) \wedge \Phi + v \wedge \mathcal{L}_u \Phi$ holds. But generally $\mathcal{L}_{fu} \neq f\mathcal{L}_u$, only for $D_u$ we have $\forall f \in E_0$: $D_{fu} = fD_u$.
3. Following the construction of the tensor bundles $\mathcal{T}^r_s(M)$ out of $T(M)$ one can associate to $V$ its dual bundle $V^*$ and the tensor products
$$V^r_s = \overbrace{V \underset{\pi}{\otimes} V \underset{\pi}{\otimes}}^{r} \cdots \overbrace{V^* \underset{\pi}{\otimes} V^* \underset{\pi}{\otimes}}^{s} \cdots V^*.$$
The differential processes $D$ and thus $\mathcal{L}_u$ can be carried over to the sections $S_p(V^r_s)$ by postulating Leibniz' rule for $p = 0$,
$$D(\Phi_1 \otimes \Phi_2) = (D\Phi_1) \otimes \Phi_2 + \Phi_1 \otimes D\Phi_2 \quad \text{for} \quad \Phi_i \in S_0(V),$$
and the invariance of the map
$$(S_0(V^*) \times S_0(V)) \to E_0(B):$$
$$D(\Psi|\Phi) = (D\Psi|\Phi) + (\Psi|D\Phi), \qquad \Psi \in S_0(V^*), \qquad \Phi \in S_0(V).$$
$D(\Psi|\Phi) = d(\Psi|\Phi)$ because $D = d$ on the scalars. In particular, for the dual basis $b^{*j}$, $(b^{*j}|b_k) = \delta^j{}_k$ we infer from $d\delta^j{}_k = 0$ that $Db^{*j} = -\omega^j{}_k b^{*k}$. One typical example will be $S(L(F))$ where $L(F)$ is identified with $F^* \otimes F$: If $L(F) \ni L = \sum v_i w_i$, $v_i \in F$, $w_i \in F^*$ then $L \cdot u = \sum_i v_i(w_i|u)$. Again $\mathcal{L}_u$ depends only on the direction of $u$ and not on its derivative when applied to $S_0(V^r_s)$ where it is denoted by $D_u$.

**Geometrical Significance of the Covariant Derivative** (4.1.17)

A section $\Phi \in S_0(V)$ as map $B \to V$ also maps $T(B) \xrightarrow{T(\Phi)} T(V)$. With reference to a basis this can be written: $x \to (x, \varphi^i(x))$,
$$T(\Phi)\colon (x, v) \to \left(x, \varphi^i(x); v, v^\alpha \frac{\partial \varphi^i}{\partial x^\alpha}\right).$$

If we identify $F$ and $T(F)$ this looks schematically as follows.



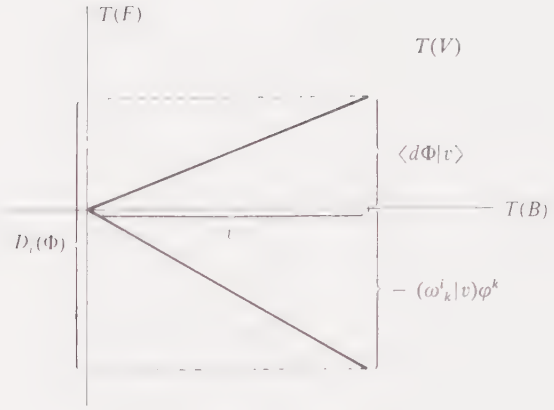

What we now want for $D_v(\Phi)$ is the change of the vertical part of $\Phi$ (in the direction of $F$) as one goes in the direction $v$ of $B$. However, in $T(V)$ there is no preferred horizontal direction, only the vertical direction is defined by $T(\Pi)$. For the components in the vertical direction we need a preferred horizontal direction. The latter is defined by the 1-forms $-\omega^i{}_k$ which can be considered as map†

$$(x, v) \to (x, \varphi^i(x); v, -(\omega^i{}_k | v)\varphi^k).$$

$D_v(\Phi)$ as difference between the two maps measures the change $T(\Phi)v$ relative to the horizontal direction given by $\omega$ (see the following illustration). Thus $D_v\Phi$ should be looked at as the change of $\Phi_x \in F$ as one proceeds in direction of $v$. It comes about because the components change by $(d\varphi^i | v)$ and the basis is rotated by $(\omega^i{}_k | v)$.

A section whose vertical component does not change will be one with $D_v\Phi = 0\ \forall v$, that is $D\Phi = 0$. One says that such a section is covariant constant or $\Phi_x \in F$ is parallel transported as one moves in $B$. If one has a covariant constant basis the $\omega$'s are zero and the covariant constant vectors are the ones with constant components. In this case the notion of parallelity given infinitesimally by $D$ can be extended to the domain of this basis. One can see on $T(S^2)$ that in general the extension even to the local level might be impossible. There the standard prescription for parallel transport along great circles is to keep the angle with them fixed. If one now takes a vector at the north pole and transports it to the south pole one arrives at vectors pointing in different directions if one takes different great circles to get there. Even for points which are a small but finite distance apart the above description does not lead to a unique result. Obviously, a necessary condition for $D\Phi = 0$ is $DD\Phi = 0$. Unlike $d$ the square of $D$ is not identically zero. It

† Analogously to $\times_\Pi$ the sum $+_\Pi$ of vector bundles is defined such that the fibres are the sum of the fibres and the basis is the common basis.




The geometrical meaning of $D_v$

turns out that $D \cdot D$ does not depend on the derivatives of the section one applies it to, our rules (4.1.8) say

$$\begin{aligned} DD(f \wedge \Phi) &= D(df \wedge \Phi + (-)^q f \wedge D\Phi) \\ &= ddf \wedge \Phi + (-)^{q+1} df \wedge D\Phi + (-)^q df \wedge D\Phi \\ &\quad + f \wedge DD\Phi = f \wedge DD\Phi \qquad \forall f \in E_q, \quad \Phi \in S_p. \end{aligned} \tag{4.1.18}$$

Thus $DD\Phi \in S_{p+2}$ can be linearly expressed by $\Phi$ and thus defines a 2-form with values in $L(F)$.

**Definition** (4.1.19)

The **curvature form** $\Omega \in S_2(L(F))$ is defined by

$$DD\Phi = \Omega \wedge \Phi, \qquad \Phi \in S_p.$$

**Remarks** (4.1.20)

1. $\Omega$ is determined by its action on a basis $\Omega b_i \equiv b_k \Omega^k{}_i$ since (4.1.18) shows $DDb_i \varphi^i = b_k \Omega^k{}_i \varphi^i$. Now

$$DDb_i = D(b_k \omega^k{}_i) = b_k \, d\omega^k{}_i + b_k \omega^k{}_j \wedge \omega^j{}_i$$

and we arrive at "Cartan's second structure equation"

$$\Omega^i{}_k = d\omega^i{}_k + \omega^i{}_j \wedge \omega^i{}_k.$$

(In the tangent bundle of space-time we shall use the letter $R$ for $\Omega$.) Considering $\omega \in S_1(L(F))$ one may write this more compactly $\Omega = d\omega + \omega \wedge \omega$. Since matrix multiplication is not commutative $\omega \wedge v \neq -v \wedge \omega$ for matrix valued 1-forms and $\omega \wedge \omega$ need not vanish.

2. For the connection (4.1.13) in $V = \mathbb{R}^4 \times \mathbb{R}^2$ with the abelian gauge group $\mathscr{A}(2)$ the term $(\omega \wedge \omega)^i{}_k = A \wedge A \varepsilon^i{}_j \varepsilon^j{}_k$ vanishes since $A \wedge A = 0$

and $\Omega^i{}_k = dA\varepsilon^i{}_k = F\varepsilon^i{}_k$ ($\varepsilon^i{}_k = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$). Thus the electromagnetic field $F = dA$ is the curvature of this connection.

3. By a change of basis $b_i \to \bar{b}_i = b_k L^k{}_i$ we have because of (4.1.18)

$$DD\bar{b}_i = b_j \Omega^j{}_k L^k{}_i = \bar{b}_k \Omega^k{}_i,$$

which shows that

$$\bar{\Omega}^i{}_k = (L^{-1})^i{}_j \Omega^j{}_m L^m{}_k.$$

Thus $\Omega$, in contradistinction to $\omega$, transforms under a change of basis like an element of $S_2(L(F))$ ought to transform. One can verify (Problem 4) that the inhomogeneous terms in the transformation law (4.1.12) do cancel out in the combination $d\omega + \omega \wedge \omega$.

4. To see the significance of $\Omega$ for the directional derivative we have to use the identity for $\Phi \in S_0$

$$(i_v i_u DD)\Phi = i_u D i_v D\Phi - i_v D i_u D\Phi - D_{[u,v]}\Phi \tag{4.1.21}$$

(Problem 7). It tells us that if we have two vector fields with $[u, v] = 0$ so that they form two-dimensional surfaces then $(D_u D_v - D_v D_u)\Phi = (i_v i_u \Omega)\Phi$. We see that $\Omega$ measures the difference between the covariant changes of $\Phi$ by first proceeding in the direction of $u$ and then in the direction of $v$ and secondly by coming to the same (infinitesimally adjacent) point by proceeding in the opposite order.

We return to the question of covariant constant sections. $D\Phi = 0$ means for the components of $\Phi$

$$d\varphi^i + \omega^i{}_k \varphi^k = 0. \tag{4.1.22}$$

This implies

$$0 = d(\omega^i{}_k \varphi^k) = (d\omega^i{}_k + \omega^i{}_j \wedge \omega^i{}_k)\varphi^k = \Omega^i{}_k \varphi^k. \tag{4.1.23}$$

Thus a covariant constant section has to be at each point $x \in B$ an eigenvector of the matrix $\Omega^i{}_k(x)$ with eigenvalue zero. If we have a covariant constant basis $\Omega$ has to be zero since its action on a basis gives zero. This also follows from the fact that in such a basis $\omega = 0$ and therefore $\Omega$ is zero and $\Omega$ transforms homogeneously under a change of basis. Conversely, $\Omega = 0$ implies $d\omega^i{}_k + \omega^i{}_j \wedge \omega^j{}_k = 0$ which are the integrability condition for the system (4.1.22). Frobenius' theorem [22, Vol. I] tells us that they are also sufficient for the local existence of solutions of (4.1.22) with arbitrary initial conditions. This situation deserves a name.

**Definition** (4.1.24)

A connection of a vector bundle is called locally **flat** if one of the two equivalent conditions hold:

(i) $\Omega = 0$;
(ii) there exists (locally) a covariant constant basis.

**Remarks** (4.1.25)

1. If the bundle is parallelizable, that is, there is a global basis, then we get a flat connection by declaring this basis as covariant constant. However, flatness is a local property and a flat bundle like the Möbius strip need not be a product. But even if the tangent space of a manifold is a flat product bundle the manifold need not be an open set of $\mathbb{R}^n$ but can have different global properties as is the case for $T^n$.
2. Flatness is an internal property of the manifold. $\Omega$ does not reproduce the curvature of a surface as imbedded in, say, $\mathbb{R}^3$. $\Omega$ vanishes for a cylinder because it could be rolled out flat on a plane without affecting its metric and thereby its internal structure.

The covariant exterior derivative $D$ can be extended in a natural way to $S_p(L(F))$ (see (4.1.16; 3) and one finds when applied to $\Omega$.

**Bianchi's Identity** (4.1.26)

$$D\Omega = 0.$$

**Remarks** (4.1.27)

1. For the components (4.1.20) of $\Omega = b_k \Omega^k{}_i b^{*i}$ we see from (4.1.16; 3)
$$0 = D\Omega = b_r(\omega^r{}_k \Omega^k{}_j + d\Omega^r{}_j - \Omega^r{}_i \omega^i{}_j)b^{*j}$$
or
$$d\Omega^i{}_j = -\omega^i{}_k \wedge \Omega^k{}_j + \Omega^i{}_k \wedge \omega^k{}_j.$$
This relation is proved in Problem 6.
2. One might think $\Omega = D\omega$ and thus $D^2 = 0$ on $\omega$. However, it has to be noted that though $\Omega$ is a covariant exterior derivative of $\omega$ in the sense that $\Omega$ transforms like $L(F)$ under a change of basis this $D$ is not the one used in (4.1.26) since $\omega$ does not transform like $S_1(L(F))$. Written in components as above we would get
$$(D\omega)^i{}_k = d\omega^i{}_k + 2\omega^i{}_l \wedge \omega^l{}_k$$
and differs from $\Omega^i{}_k$ by the factor 2 in the last term.

Our general considerations so far do not constrain $\omega$ any further. We shall now impose restrictions on the covariant derivative which finally determine it uniquely from the fibre metric.

The fibre metric defines a scalar product in $F$ and thus maps $S_p(V) \times S_0(V)$ into $E_p(B)$. We shall require that the connection conserves this structure in the sense that the derivative of the scalar product contains just the derivatives of the sections involved.

**Compatibility of the Connection with the Fibre Metric** (4.1.28)

In a vector bundle $V$ with fibre metric $\langle\,|\,\rangle$ we shall require

$$D\langle\Psi|\Phi\rangle = d\langle\Psi|\Phi\rangle = \langle D\Psi|\Phi\rangle + (-)^p\langle\Psi|D\Phi\rangle,$$
$$\Psi \in S_p(V), \qquad \Phi \in S_0(V).$$

Such connections will be called metric connections.

**Remark** (4.1.29)

For the $\omega$'s this imposes the condition

$$\begin{aligned} 0 = d\eta_{ik} = d\langle e_i|e_k\rangle &= \langle De_i|e_k\rangle + \langle e_i|De_k\rangle \\ &= \omega^j{}_i\langle e_j|e_k\rangle + \langle e_i|e_j\rangle\omega^j{}_k = \omega^j{}_i\eta_{jk} + \omega^j{}_k\eta_{ij}. \end{aligned}$$

Thus if the $\eta$ are symmetric $\omega_{ik} \equiv \omega^j{}_k\eta_{ij}$ has to be antisymmetric

$$\omega_{ik} = -\omega_{ki}. \tag{4.1.30}$$

This means that $\omega_{ik}(x)$ cannot be an arbitrary element of $L(F)$, but must belong to the Lie algebra of the gauge group. In our cases $G$ will be for the inner gauge symmetries an $O(n)$ or $U(n)$ group and for the space-time symmetry we have $O(3, 1)$. For an arbitrary basis with $\langle e_i|e_k\rangle = g_{ik}$ (4.1.30) generalizes to $dg_{ik} = \omega_{ik} + \omega_{ki}$.

**Examples** (4.1.31)

1. The Lie algebra of the one-dimensional group $O(2)$ are the multiples of $\left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$ and the connection (4.1.14; 2) conserves the $O(2)$-invariant fibre metric $\langle\Phi|\Phi\rangle = (\varphi^1)^2 + (\varphi^2)^2$ since $\omega^i{}_k = \omega_{ik} = -\omega_{ki}$.
2. In the tangent bundle of a Riemannian space $\eta_{ik} = \delta_{ik}$ and thus the $\omega^i{}_k = -\omega^k{}_i$. For the pseudo-Riemannian space-time we have

$$\eta_{ik} = \begin{cases} -1 & \text{if } i = k = 0, \\ 1 & \text{if } i = k = 1, 2, 3, \\ 0 & \text{otherwise.} \end{cases}$$

Hence $\omega^i{}_k = -\omega^k{}_i$ for $i, k = 1, 2, 3$, $\omega^i{}_0 = \omega^0{}_i$, and $\omega^i{}_i = 0$.

We shall finally study special properties of the tangent bundle of a Riemannian space. To indicate that we are considering this special case we shall denote the curvature in $T(M)$ by the traditional letter $R$. First, we use the fact that in this case the sections $S_0(T(M))$ can be naturally identified with the vector fields $\mathcal{T}^1_0(M)$. Since vector fields $\mathcal{T}^1_0(M)$ map $E_1$ into $E_0$ they also map $S_1(T(M))$ into $S_0(T(M)) \equiv \mathcal{T}^1_0(M)$. The vector valued 1-form which produces in this way the identity map has a special name.

**Definition** (4.1.32)

If the $b_i$ form a local basis in $\mathcal{T}_0^1(M)$ and $e^i$ the dual basis, $(e^i|b_j) = \delta^i{}_j$, then

$$\theta \equiv \sum_i b_i \otimes e^i \in S_1(T(M))$$

is called **soldering form** and $T \equiv D\theta$ **torsion**. Metric connections with $T = 0$ are called torsion-free or Levi-Civita connections.

**Remarks** (4.1.33)

1. $D\theta = \sum b_i(de^i + \omega^i{}_k \wedge e^k)$ and for a torsion-free connection we have $de^i = -\omega^i{}_k \wedge e^k$. Unless specified otherwise we will only consider torsion-free metric connections.
2. In Problem 11 it is verified that $T = 0$ is equivalent to $D_X Y - D_Y X = [X, Y]$, $X, Y \in \mathcal{T}_0^1$ or to $(D_X u|Y) - (D_Y u|X) = (du|X \otimes Y)$, $u \in \mathcal{T}_1^0$.

**Examples** (4.1.34)

1. In a flat Riemannian space we have in the covariant constant basis $\omega^i{}_k = 0$. If, in addition, $T = 0$ we see from (4.1.31; 1) that $de^i = 0$ or locally $e^i = dx^i$. Therefore $R = T = 0$ imply the local existence of a natural covariant constant basis in $E_1$.
2. Consider polar coordinates in $\mathbb{R}^2 \setminus \{0\}$. The metric is $g = dr^2 + r^2\, d\varphi^2 = e^1 \otimes e^1 + e^2 \otimes e^2$ in the orthogonal basis $e^1 = dr$, $e^2 = r\, d\varphi$. From $de^1 = 0$, $de^2 = dr \wedge d\varphi$ we conclude for a torsion-free connection which preserves $g$ that $\omega_{12} = -\omega_{21} = -d\varphi = -e^2/r$, $\omega_{11} = \omega_{22} = 0$. Thus $R_{ik} = d\omega_{ik} + \omega_i{}^l \wedge \omega_{lk} = 0$. With a connection for which the orthogonal

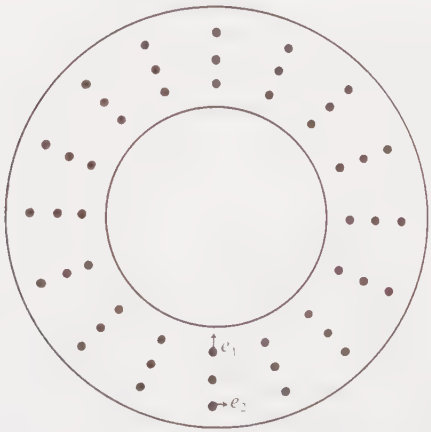



Figure 54 The bent crystal, a model of a space with torsion

basis is covariant constant we get the torsion $b_2\, de^2 = (1/r)\, \partial\varphi \otimes dr \wedge d\varphi$ since the dual basis is $b_1 = \partial r$, $b_2 = (1/r)\, \partial\varphi$. One verifies that in this case the relation $D_X Y - D_Y X = [X, Y]$ from (4.1.31; 2) is violated. If we take $X = b_1$, $Y = b_2$ the left-hand side vanishes since the basis is covariant constant but $[b_1, b_2] = [\partial r, 1/r]\, \partial\varphi = -(1/r^2)\, \partial\varphi \neq 0$. This definition of the connection could be made concrete by bending a rectangular crystal into a ring so that the atoms sit at the points $r = nr_0$, $\varphi = m\varphi_0$ for $n, m \in \mathcal{N}$ (Figure 54). If we lived in such a crystal and defined parallel transport by the lattice of atoms as usually, then our space would have torsion.

3. The sphere $S^2$ with the metric $g = R^2(d\theta^2 + \sin^2\theta\, d\varphi^2)$. In the orthogonal basis $e^1 = R\, d\theta$, $e^2 = R \sin\theta\, d\varphi$ we have $de^1 = 0$, $de^2 = R\cos\theta\, d\theta \wedge d\varphi$ and therefore $\omega_{12} = -\omega_{21} = -\cos\theta\, d\varphi$. This gives a curvature

$$R^1{}_2 = d\omega^1{}_2 = \sin\theta\, d\theta \wedge d\varphi = \frac{1}{R^2}\, e^1 \wedge e^2.$$

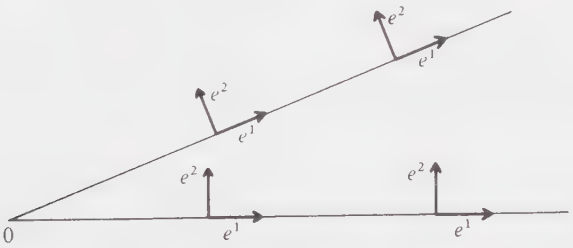
4. The Lorentz hyperboloid $H = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 - z^2 = R^2\}$ and $g$ = restriction of $dx^2 + dy^2 - dz^2$ to $H$. In cylindrical coordinates $(x, y) = \rho(\cos\varphi, \sin\varphi)$ we find that $e^1 = R\, d\rho/\sqrt{\rho^2 - R^2}$ and $e^2 = \rho\, d\varphi$ are an orthogonal basis on $H$: $g_{|H} = -e^1 \times e^1 + e^2 \times e^2$. $de^1 = 0$, $de^2 = d\rho \wedge d\varphi \Rightarrow \omega^1{}_2 = -\omega^2{}_1 = -\sqrt{\rho^2/R^2 - 1}\, d\varphi \Rightarrow R^1{}_2 = d\omega^1{}_2 = -\rho R^{-2}/\sqrt{\rho^2/R^2 - 1}\, d\rho \wedge d\varphi = -R^{-2} e^1 \wedge e^2$.

**Remark** (4.1.35)

In Example 2, those $\omega$'s imply for the covariant derivative that

$$D_X e^2 = -(e^2 | X)\frac{e^1}{r}, \qquad D_X e^1 = (e^2 | X)\frac{e^2}{r}.$$

If $X$ has only the 1-component, then the $D_X e^i$ vanish. The covariant derivative of $e^i$ in the 2-direction is a rotation, which is faster for smaller $r$. This accords exactly with the following intuitive picture of the rotation of the local basis in polar coordinates.




The rotating basis of polar coordinates

We have seen that the two conditions of invariance of $g$ under parallel transport and absence of torsion imply the relations (1.2.25). There it was stated that they determined the connection $\omega$. We shall now verify this claim by an explicit expression of $\omega$ in terms of the derivatives of $g$.

**Calculation of the Affine Connections** (4.1.36)

(a) In the natural basis, $e^l = dq^l$: The so-called **Christoffel symbols** $\Gamma_{ijk} \in \mathcal{F}_0^0$ occur in the decomposition of $\omega_{ij}$ in the basis, $\omega_{ij} = \Gamma_{ijk}\, dq^k$.
Since in the present case $de^j = 0 = \Gamma_{ijk}\, dq^j \wedge dq^k$, they are symmetric in $j$ and $k$. In this basis, $dg_{ik} = g_{ik,l}\, dq^l$, so the $\Gamma$'s are determined by the facts

$$g_{jk,l} = \Gamma_{jkl} + \Gamma_{kjl}$$

and

$$\Gamma_{ijk} = \Gamma_{ikj},$$

which are solved by

$$\Gamma_{jkl} = \tfrac{1}{2}(g_{jk,l} + g_{lj,k} - g_{kl,j}) \tag{4.1.37}$$

Since this solution is unique, we see that our axioms fix $D$ uniquely (Problem 5).

(b) In an orthogonal basis: From $dg_{ik} = d\eta_{ik} = 0 = \omega_{ik} + \omega_{ki}$ together with

$$(de^j | e_k \otimes e_i) = -(\omega^j{}_l \wedge e^l | e_k \otimes e_i) = (\omega^j{}_k | e_i) - (\omega^j{}_i | e_k),$$

one can derive

$$(\omega_{kj} | e_i) = \tfrac{1}{2}[(de_j | e_i \otimes e_k) + (de_k | e_j \otimes e_i) - (de_i | e_k \otimes e_j)] \tag{4.1.38}$$

(Problem 5).

Next we study the special features which arise because a flow in the base manifold induces in a natural way a flow in all the tensor bundles. Thus the Lie derivative of a section with respect to a vector field in $B$ makes sense and we can study the

**Commutator of the Covariant Derivative and the Lie Derivative** (4.1.39)

We noted in (4.1.21) that for vector fields $u$, $v$, $X$ the commutator $(D_u D_v - D_v D_u - D_{[u,v]})X = (i_v i_u R)X \equiv R(u, v)X$ does not depend on the derivative of $X$ but defines the curvature. One can derive a similar relation for $DL_v - L_v D$ using (4.1.21) and the fact that the Lie derivative $L_v$ (like $D_v$) applied to the inner product of a vector with a 1-form just takes the derivative of both: $L_v i_u X = i_{L_v u} X + i_u L_v X$. (For the inner product of forms this is true only if $v$ is a Killing vector field.) Remembering $[X, Y] \equiv L_X Y = -L_Y X = D_X Y - D_Y X$ we find

$$\begin{aligned}
\langle (DL_v - L_v D)X | u \rangle &= D\, L_v X - L_v D_u X + D_{[v,u]} X \\
&= D_u D_v X - D_u D_X v - D_{[u,v]} X - D_v D_u X + D_{D_u X} v \\
&= R(u, v)X + D_{D_u X} v - (R(u, X)v + D_X D_u v + D_{[u,X]} v) \\
&= R(u, v)X + R(X, u)v + D_{D_u X} v - D_X D_u v - D_{D_u X} v + D_{D_X u} v \\
&= R(X, v)u - \langle D_X D v | u \rangle.
\end{aligned}$$

In the last step we used a cyclic property of $R$ derived in (4.1.44; 2(c)) and $D_X D_u v = D_X i_u Dv = i_{D_X u} Dv + i_u D_X Dv = D_{D_X u} v + \langle D_X Dv | u \rangle$. Thus we see that the commutator does not contain any derivatives of $X$ but second derivatives of $v$.

### Commutativity of $D$ and * (4.1.40)

In (1.2.16) we widened the scalar product $\langle\ |\ \rangle$ to the interior product $i_v$. (4.1.28) generalizes for $i_v$ to $D_X i_v = i_{D_X v} + i_v D_X$ as used above. In particular, we see $D_X \varepsilon = 0$ since $0 = D_X 1 = D_X \langle \varepsilon | \varepsilon \rangle = 2 \langle D_X \varepsilon | \varepsilon \rangle$ and $\varepsilon$ is not degenerate. Therefore $D_X i_v \varepsilon = i_{D_X v} \varepsilon$ and thus $D_X {}^* v = {}^* D_X v$. Here $E_p$ is identified with $S_0(\wedge_p T^*(B))$ and not with $S_p(T^*(B))$. Thus $D_X$ on $E_p$ cannot be written as $i_X D$ where $i_X$ is the map $E_p \to E_{p-1}$ in which sense it is to be understood here.

In a pseudo-Riemannian space vector fields define directions in the vector bundles. Thus the covariant derivative $D_X$ distinguishes the vector fields that are consistent with it in the sense that a vector translated parallelly in its own direction from some point equals the vector at the neighboring point:

### Definition (4.1.41)

A vector field $X$ such that $D_X X = 0$ is called a **geodesic vector field**.

### Remarks (4.1.42)

1. The connection with the geodesic lines: Let $z(s)$ be a streamline of a geodesic vector field $X$, that is, $\dot{z}(s) = X(z(s))$. Then the components of $z$ in the natural basis satisfy the equation $\ddot{z}^i = -\Gamma^i{}_{jk} \dot{z}^j \dot{z}^k$, because with Remark (4.1.16; 3),

$$0 = D_X(X^i \partial_i) = (X^i{}_{,k} X^k + (\omega^i{}_j | \partial_k) X^j X^k) \partial_i$$

and

$$\ddot{z}^i(s) = \frac{d}{ds} X^i(z(s)) = X^i{}_{,k} \dot{z}^k.$$

From (4.1.36(a)), $(\omega^i{}_j | \partial_k)$ is just $\Gamma^i{}_{jk}$, and we obtain exactly the geodesic equations of motion (I: 1.1.6). The shortest lines are thus the straightest, in the sense of the tangent vector to a curve being transformed into itself under parallel transport along them.

2. The connection with Killing vector fields: Remember that for a Killing vector field $X$ we had $L_X g = 0$, whereas $L_X X = 0$ for any vector field. In contradistinction for the covariant derivative $D_X X = 0$ holds only for geodesic vector fields and $D_X g = 0$ holds for any vector field. In general "Killing" does not imply "geodesic" nor vice versa. But there is the

following relation: Let $v$ be a Killing vector field (2.1.9), so that $L_v\langle X|X\rangle = 2\langle L_v X|X\rangle$, and let $X$ be a geodesic vector field. Then

$$\begin{aligned} L_X\langle v|X\rangle &= D_X\langle v|X\rangle = \langle D_X v|X\rangle = \langle D_v X|X\rangle - \langle L_v X|X\rangle \\ &= \tfrac{1}{2}(D_v\langle X|X\rangle - L_v\langle X|X\rangle) = 0. \end{aligned}$$

Consequently, the component of a geodesic vector field in the direction of a Killing vector field is constant along a geodesic line. This reveals a new significance of a Killing vector field $v$, that it provides a constant of the motion of a particle in a gravitational field, namely the component of the four-momentum $\dot{z}$ of the particle in the direction of $v$.

3. If one vector field in an orthogonal basis is also natural it is geodesic. To see this, let $X$ be this vector field and $v$ be either $X$ (then $\langle v|X\rangle = \pm 1$) or a member of a natural basis orthogonal to $X$. In any case $\langle v|X\rangle$ is constant and with (4.1.33; 2)

$$\begin{aligned} 0 &= L_X\langle v|X\rangle = D_X\langle v|X\rangle = \langle D_X v|X\rangle + \langle v|D_X X\rangle \\ &= D_v\tfrac{1}{2}\langle X|X\rangle - \langle L_v X|X\rangle + \langle v|D_X X\rangle. \end{aligned}$$

Now $\langle X|X\rangle = \pm 1$ so the first term vanishes and $L_v X = 0$ since the Lie derivatives between members of a natural basis vanish. Therefore all components of $D_X X$ are zero.

In a pseudo-Riemannian space the curvature forms are not arbitrary, but are subject to certain algebraic conditions which we now enumerate:

**Algebraic Identities Satisfied by the Curvature Forms** (4.1.43)

(a) $R_{ij} = -R_{ji}$;
(b) $R_{ij} \wedge e^j = 0$.

**Proof**

(a) This fact is independent of the basis. If the basis is changed by $\bar{e} = Ae$ (Problem 4), then $R \to ARA^{-1}$ and $g \to A^{t-1}gA^{-1}$, and thus $R_{ij} = g_{ik}R^k{}_j \equiv (gR)_{ij} \to (A^{t-1}gRA^{-1})_{ij}$, which preserves the antisymmetry. And Fact (a) is obviously true in an orthogonal basis, since $\omega_{ij} = -\omega_{ji} \Rightarrow \omega_{ik} \wedge \omega^k{}_j = -\omega_{kj} \wedge \omega_i{}^k = -\omega_{jk} \wedge \omega^k{}_i$.
(b) $0 = d\,de^i = -d(\omega^i{}_k \wedge e^k) = -R^i{}_k \wedge e^k$. □

**Consequences** (4.1.44)

1. The $m$ 3-forms (b) we get from the $\binom{m}{2}$ 2-forms vanish. In the absence of other algebraic conditions there are consequently

$$\binom{m}{2}^2 - m\binom{m}{3} = \frac{m^2(m^2-1)}{12}$$

independent components of the curvature. In one dimension, there is no curvature 2-form, in two dimensions there is one, and in four dimensions there are twenty. However, many of the components can be made to vanish by a suitable choice of bases and coordinate systems. Let us take an orthogonal basis, to fix the $g_{ik}$ as $\eta_{ik}$, and try to find how many invariants can be constructed from the $m^2(m^2-1)/12$ components of $R_{ik}$ and the $m^2$ components of the $e^i$. There are available $m^2$ functions $\partial\bar{x}^i/\partial x^j$ from a possible change of coordinate system and $m(m-1)/2$ components of $A$ (since $A\eta A^T = \eta$), with which to make $m^2 + m(m-1)/2$ of the total of $m^2 + m^2(m^2-1)/12$ components vanish. There remain $m^2(m^2-1)/12 - m(m-1)/2 = m(m-1)(m-2)(m+3)/12$ invariants. This simple argument does not work in two dimensions, however. Moreover, we have only imposed the algebraic conditions, and there may be some differential dependence among the invariants as well.

2. If $R^i{}_j$ are decomposed in a basis $R^i{}_j = \frac{1}{2}R^i{}_{jkm}e^{km}$, then the **Riemann–Christoffel tensor** $R_{ijkm} = g_{in}R^n{}_{jkm}$ obeys the equations (Problem 8)

   (a) $R_{ijkm} = -R_{ijmk}$;
   (b) $R_{ijkm} = -R_{jikm}$;
   (c) $R_{ijkm} + R_{ikmj} + R_{imjk} = 0$;
   (d) $R_{ijkm} = R_{kmij}$.

3. If $R_{jk}$ is written in terms of the contractions $i_{e_j}R^j{}_k = R^j{}_{kjm}e^m \equiv R_k \in E_1$ and $i_{e_k}R^k = R^{jk}{}_{jk} \equiv R \in E_0$ in such a way that the contractions of the remainders vanish, then those remainders define the **Weyl forms** $C_{jk}$:

$$R_{jk} \equiv \frac{-R}{(m-2)(m-1)}e_j \wedge e_k + \frac{1}{m-2}(e_j \wedge R_k - e_k \wedge R_j) + C_{jk}.$$

It is clear that the $C_{ij}$ also satisfy (4.1.23), and by construction $i_{e_j}C^j{}_k \equiv C_k = 0$ (Problem 9). Because $i_{e_j}R_k = i_{e_k}R_j$, the equations $C_k = 0$ pose only $m(m+1)/2$ independent conditions, which leaves $m^2(m^2-1)/12 - m(m+1)/2 = m(m+1)(m+2)(m-3)/12$ components for the $C_{ij}$. In three dimensions all $C_{ij}$ are zero; they first occur in four dimensions, with ten components. The $C^i{}_j$ are important because they remain invariant under conformal transformations $g \to fg, f \in E_0$ (Problem 10), and in particular they vanish on all conformally flat spaces. Conversely, if $C = 0$, then $g = f\eta$ [25].

We conclude the section by collecting the formulas that connect the metric and the curvature, to lay the foundation for later calculations:

$$\begin{aligned} g &= g_{ik}e^i \otimes e^k, \\ de^i &= -\omega^i{}_j \wedge e^j, \\ dg_{ik} &= g_{ij}\omega^j{}_k + \omega_k{}^j g_{ji}, \\ R^i{}_k &= d\omega^i{}_k + \omega^i{}_j \wedge \omega^j{}_k. \end{aligned} \tag{4.1.45}$$

**Problems** (4.1.46)

1. Find a $u$ such that $(du|X \otimes Y) = (D_X u|Y) - (D_Y u|X)$ is violated in Example (4.1.34; 2) with torsion.
2. Use (1.2.26; 3) to calculate the $\omega$'s for the plane with polar coordinates. (The $\omega$'s vanish for $dx^{1,2}$, so set $\bar{e}^i = A^i{}_k\, dx^k$, and compare with (4.1.34; 2).)
3. With the help of (1.2.26; 3), find a necessary and sufficient condition to make the $\omega$'s vanish by a change of basis.
4. Show that for $\bar{b} = bL$, we have $\Omega = L^{-1}\Omega L$.
5. Derive (4.1.37) and (4.1.38).
6. Prove (4.1.27; 1).
7. Show (4.1.21) by verifying the more general relation for $\omega \in S_1$:

$$i_v i_u D\omega = i_u D(\omega|v) - i_v D(\omega|u) - i_{[u,v]}\omega.$$

   *Hint*: $\omega$ can be written $\sum_i b_i v^i, b_i \in S_0$, $v^i \in E_1$ and because of linearity it suffices to show the relation for one term,. Use

$$i_v i_u dv = i_u d(v|v) - i_v d(v|u) - i_{[u,v]}v \qquad \text{for } v \in E_1.$$

8. Derive the equations (4.1.44; 2).
9. Show that $i_{e_j} C^j{}_k \equiv C_k = 0$.
10. Suppose that $\bar{e}^i = fe^i$ but $\bar{g}_{ik} = g_{ik}$ such that $g = f^2 g$. Show that $\bar{C}_{ik} = C_{ik}$.
11. Show $de^j = -\omega^j{}_k e^k \overset{a}{\Leftrightarrow} (D_X u|Y) - (D_Y u|X) = (du|X \otimes Y) \overset{b}{\Leftrightarrow} D_X Y - D_Y X = [X, Y]$, $X, Y \in \mathcal{T}_0^1$, $u \in \mathcal{T}_1^0$ and $e^j$ a basis in $\mathcal{T}_1^0$.

**Solutions** (4.1.47)

1. $u = e^2$; $de^2 = dr \wedge d\varphi \neq 0$, but $D_X e^2$ ought to vanish for all $X \in T_0^1$.
2. $A = \begin{vmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{vmatrix}$, $dA = d\varphi \begin{vmatrix} -\sin\varphi & \cos\varphi \\ -\cos\varphi & -\sin\varphi \end{vmatrix}$, $-dA \cdot A^{-1} = d\varphi \begin{vmatrix} 0 & -1 \\ 1 & 0 \end{vmatrix}$.
3. Choose $A$ so that $dA^i{}_k = \omega^i{}_l A^l{}_k$. As remarked in (4.1.20; 3), it is both necessary and sufficient for solubility that $d(\omega^i{}_l A^l{}_k) = R^i{}_l A^l{}_k = 0$, and therefore $R^i{}_k = 0 \Leftrightarrow \omega^i{}_k = 0$.
4. $\bar{\omega} = L^{-1}\omega L + L^{-1}\, dL$,

$$\begin{aligned} d\bar{\omega} + \bar{\omega} \wedge \bar{\omega} &= (dL^{-1}) \wedge \omega L + L^{-1}\, d\omega L - L^{-1}\omega\, dL + (dL^{-1}) \wedge dL \\ &\quad + (L^{-1}\omega L + L^{-1}\, dL) \wedge (L^{-1}\omega L + L^{-1}\, dL) \\ &= L^{-1}(d\omega + \omega \wedge \omega)L. \end{aligned}$$

5. $$\begin{aligned} \Gamma_{kjl} + \Gamma_{jkl} &= g_{jk,l} \\ \Gamma_{jlk} + \Gamma_{ljk} &= g_{lj,k} \\ -\Gamma_{lkj} - \Gamma_{klj} &= -g_{kl,j} \\ \hline 2\Gamma_{jkl} &= g_{jk,l} + g_{lj,k} - g_{kl,j}. \end{aligned}$$

Let $(de_j|e_i \otimes e_k) = (de_j)_{ik}$, etc. Then

$$\begin{aligned} (de_j)_{ik} &= (\omega_{ji})_k - (\omega_{jk})_i \\ (de_k)_{ji} &= (\omega_{kj})_i - (\omega_{ki})_j \\ -(de_i)_{kj} &= -(\omega_{ik})_j + (\omega_{ij})_k \\ \hline (de_j)_{ik} + (de_k)_{ji} - (de_i)_{kj} &= 2(\omega_{kj})_i. \end{aligned}$$

6. $$\begin{aligned} d(d\omega^i{}_j + \omega^i{}_s \wedge \omega^s{}_j) &= d\omega^i{}_s \wedge \omega^s{}_j - \omega^i{}_s \wedge d\omega^s{}_j \\ &= \Omega^i{}_s \wedge \omega^s{}_j - \omega^i{}_s \wedge \Omega^s{}_j - \omega^i{}_k \wedge \omega^k{}_s \wedge \omega^s{}_j + \omega^i{}_s \wedge \omega^s{}_k \wedge \omega^k{}_j \\ &= \Omega^i{}_s \wedge \omega^s{}_j - \omega^i{}_s \wedge \Omega^s{}_j. \end{aligned}$$

7. Insert $\omega = bv$, $b \in S_0$, $v \in E_1$

$$\begin{aligned} i_v i_u D\omega = &(i_u Db)(i_v v) - (i_v Db)(i_u v) \\ &+ b(i_u d(i_v v) - i_v d(i_u v) - i_{[u,v]} v). \end{aligned}$$

This equals

$$i_u D(\omega|v) - i_v D(\omega|u) - i_{[u,v]}\omega = i_u D(bi_v v) - i_v D(bi_u v) - bi_{[u,v]} v.$$

8. (a) holds because both sides are components of a 2-form;
   (b) follows from (4.1.23(a));
   (c) $0 = R^i{}_j \wedge e^j = R^i{}_{jkm} e^{jkm}$. Because of (a), only the cyclic permutation remains in the sum:
   (d) follows from (a) – (c):

$$\begin{aligned} R_{ijkm} &= -R_{ikmj} - R_{imjk} = R_{kimj} + R_{mijk} \\ &= -R_{kmji} - R_{kjim} - R_{mjki} - R_{mkij} \\ &= 2R_{kmij} + R_{jkim} + R_{jmki} = 2R_{kmij} - R_{jimk} \Rightarrow 2R_{ijkm} = 2R_{kmij}. \end{aligned}$$

9. Because $i_{e_j} e^j = m$ and $e^j i_{e_j} \omega = p\omega$ for all $\omega \in E_p$, we find

$$R_k = i_{e_j} R^j{}_k = \frac{-R(m-1)e_k}{(m-2)(m-1)} + \frac{1}{m-2}[Re_k - R_k - R_k + mR_k] + C_k \Rightarrow C_k = 0.$$

10. It suffices to show that $\bar{R}_{ik} = R_{ik} + v_i \wedge e_k - v_k \wedge e_i$ for some $v_i \in E_1$. Because $d\bar{e}^i = df \wedge e^i + f\, de^i = -\bar{\omega}^i{}_k \wedge \bar{e}^k$ and $\bar{\omega}_{ik} + \bar{\omega}_{ki} = \omega_{ik} + \omega_{ki}$,

$$\bar{\omega}_{ik} = \omega_{ik} + (df|e_i)e_k - (df|e_k)e_i.$$

In a natural basis, $e_s \wedge \omega^s{}_n = 0$, and therefore

$$((df|e_i)e_s - (df|e_s)e_i) \wedge \omega^s{}_k = v_k \wedge e_i, \qquad v_k = (df|e_s)\omega^s{}_k,$$

and

$$d\bar{\omega}_{ik} - d\omega_{ik} = v_i \wedge e_k - v_k \wedge e_i, \qquad v_i = d(df|e_i).$$

11. (a) $\Leftarrow$: $(de^j|X \otimes Y) = (D_X e^j|Y) - (D_Y e^j|X)$

$$= -(\omega^j{}_k|X)(e^k|Y) + (\omega^j{}_k|Y)(e^k|X) = -(\omega^j{}_k \wedge e^k|X \otimes Y)$$

$\Rightarrow$ If $u = fe, f \in E_0$ then $du = df \wedge e + f\, de$ and $D_X u = i_X(df \wedge e + f\, De)$.

Thus the first term also satisfies the equation.

(b) Remember $(du|X \otimes Y) = L_X(u|Y) - L_Y(u|X) - (u|[X, Y])$ and $L_X(u|Y) = D_X(u|Y) = (D_X u|Y) + (u|D_X Y)$.

## 4.2 Gauge Theories and Gravitation

*Energy and momentum imprint a structure on space and time through Einstein's equations, which equate the energy and momentum forms to quantities constructed from $R^{\alpha}{}_{\beta}$.*

To derive the field equations with the Lagrangian formalism one needs a 4-form as Lagrangian density. However, to get fields equations which are consistent one has to make sure that the action functional has a stationary point. We have seen in §2.1 that in Maxwell's theory this is only the case if the current is conserved. In §3.1 the current was derived from a scalar field $S$ and the field equations for $S$ were just the requirement that the current be conserved. We shall now study this phenomenon more generally with a scalar field $\Phi = \sqrt{\rho}\, e^{iS}$ where the density $\rho$ is no longer constant. Instead of the complex field $\varphi$ we shall use two real scalar fields $\varphi^1$, $\varphi^2$ with $\Phi^1 = \varphi^1 + i\varphi^2$. Expressed in terms of these fields the Lagrangian $\overline{(d + ieA)\varphi} \wedge *(d + ieA)\varphi$ of (3.1.21; 2) becomes in the notation (4.1.14; 2)

$$-\frac{1}{2}\sum_{i=1}^{2}(D\varphi)^i \wedge *(D\varphi)^i, \qquad (D\varphi)^i = d\varphi^i + e\tau^i{}_k A\varphi^k, \qquad \varepsilon = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{4.2.1}$$

Another term invariant under the local gauge transformation would be a "mass-term"$-\frac{1}{2}m^2 \sum_i \varphi^i \wedge *\varphi^i$. The general total Lagrangian

$$\mathscr{L} = -\tfrac{1}{2}\, dA \wedge *dA - \tfrac{1}{2}\sum_i ((D\varphi)^i \wedge *(D\varphi)^i + m^2\varphi^i \wedge *\varphi^i) \tag{4.2.2}$$

is invariant under $A \to A + d\Lambda$, $\varphi \to e^{-e\varepsilon\Lambda}\varphi$, $\Lambda \in C(\mathbb{R}^4)$ and the linear dependence of the action $W = \int \mathscr{L}$ on $\Lambda$ which prevented $W$ in (2.1.3; 1) to have stationary points is absent. Thus there is hope that the Euler equations which result from (4.2.2)

$$d*F = -*J, \qquad D*D\Phi = m^2 *\Phi, \tag{4.2.3}$$

with

$$*J = -\frac{\partial \mathscr{L}}{\partial A} = e\varepsilon^j{}_k \varphi^k \wedge *(D\Phi)^j$$

are consistent. Indeed, $d*J = 0$ is a consequence of the equations for $\Phi$. To see this most directly consider the variation of $\mathscr{L}$ with $\Phi$

$$\delta\mathscr{L} = \sum_i \delta\varphi^i\left[\frac{\partial\mathscr{L}}{\partial\varphi^i} - d\,\frac{\partial\mathscr{L}}{\partial\, d\varphi^i}\right] + d\left(\sum_i \delta\varphi^i\,\frac{\partial\mathscr{L}}{\partial\, d\varphi^i}\right). \tag{4.2.4}$$

The Euler equations require $[\ ] = 0$ and for $\delta\Phi$ the change $\Phi \to e^{-e\varepsilon\Lambda}\Phi$ to first order in $\Lambda$ gives $\delta\mathscr{L} = ed(\Lambda\varepsilon^j{}_k\varphi^k \wedge *(D\Phi)^j)$. If we choose a global

gauge transformation, ($d\Lambda = 0$), the invariance of $\mathscr{L}$ guarantees $\delta\mathscr{L} = 0 = -d^*J$.

Next we shall extend these considerations to the non-abelian gauge group $O(3)$. In this case $\Phi$ has three components and one might think that in (4.2.2) one just has to take $\sum_i$ from one to three. However, now one also has three gauge potentials and $dA$ is not invariant under the gauge transformation $A \to L^{-1}AL + L^{-1}dL$ (see 4.1.12). This suggests replacing $dA$ with $DA = \Omega$ (see 4.1.20; 1) which transforms as $(DA) \to L^{-1}AL$. Thus the bilinear expression $\operatorname{tr}\Omega \wedge {}^*\Omega = \Omega^i{}_k \wedge {}^*\Omega^k_i$ is gauge invariant and seems to be a good candidate for the Lagrangian of the gauge field. Such a Lagrangian has first been proposed by Yang and Mills and has later turned out to describe the strong and electroweak interactions if further fields are added.

**The Yang-Mills Lagrangian** (4.2.5)

$$\mathscr{L} = \tfrac{1}{2}\operatorname{tr} DA \wedge {}^*DA - \tfrac{1}{2}\langle D\varphi \wedge {}^*D\varphi\rangle - \tfrac{1}{2}m^2\langle\Phi \wedge {}^*\Phi\rangle.$$

**Remarks** (4.2.6)

1. We used the notation $\langle\Phi \wedge \Psi\rangle = \sum_i \varphi^i \wedge \psi^i$ for the scalar product in the fibres.
2. Although we were thinking of $O(3)$ the Lagrangian (4.2.5) has the same appearance for all $O(n)$.

The Euler equations derived from (4.2.5) are similar to (4.2.3) except that now the Maxwell field $F$ is replaced by the curvature $\Omega$. Here we just state the results as we shall go in (4.2.15) through the details of the variational procedure.

**The Yang-Mills Equations** (4.2.7)

$$d^*\Omega = -{}^*J \equiv \frac{\partial\mathscr{L}}{\partial A}, \qquad D^*D\Phi = m^2{}^*\Phi.$$

**Remarks** (4.2.8)

1. The derivation of (4.2.7) follows the lines of (2.1.7), one just notes ${}^*\Omega = \partial\mathscr{L}/\partial\, dA$ since both factors in $\Omega \wedge {}^*\Omega = {}^*\Omega \wedge \Omega$ give the same contribution.
2. To get more explicit expressions one might use a basis $b_i$ for the Lie-algebra $= (3 \times 3$ antisymmetric matrices) which diagonalizes the trace. If $\operatorname{tr} b_\alpha b_\beta = -\delta_{\alpha\beta}$ and $A = \sum b_\alpha A^\alpha$, $A^\alpha \in E_1$, then the curvature $\Omega = \sum b_\alpha F^\alpha$ has according to (4.1.20; 1) the components

$$F^\alpha = dA^\alpha + \tfrac{1}{2}c_{\beta\gamma}{}^\alpha A^\beta \wedge A^\gamma.$$

For $O(n)$ $\alpha$ goes from 1 to $n(n-1)/2$ whereas $i$ for $\varphi^i$ goes from 1 to $n$. $c_{\beta\gamma}{}^{\alpha}$ are the structure constants of $O(3)$ of this basis: $b_\beta b_\gamma - b_\gamma b_\beta = c_{\beta\gamma}{}^{\alpha} b_\alpha$. With this decomposition the Lagrangian of the gauge field becomes $\mathscr{L}^{\mathrm{gf}} = -\frac{1}{2} F^\alpha \wedge {}^*F^\alpha$.

3. Since the $F^\alpha$ do not depend on $dA$ only, but also on $A$, there is a contribution $-\partial \mathscr{L}^{\mathrm{gf}}/\partial A^\alpha = c_{\alpha\beta}{}^{\gamma} A^\beta \wedge {}^*F^\gamma$ to the current ${}^*J$. This does not appear in Maxwell's theory and means that the Yang–Mills equations are nonlinear even in the absence of the scalar field. On the other hand, the contribution of the scalar field to the current $(b^\alpha)_{kl}\varphi^k \wedge {}^*(D\Phi)^l$ has the same structure as previously.
4. The consistency of the Yang Mills equations requires $d{}^*J = 0$. This follows indeed from the variational principle as can be seen by considerations similar to the preceding ones. A gauge transformation $L = 1 + b_\alpha \Lambda^\alpha$ with $d\Lambda^\alpha = 0$ and $\Lambda^\alpha \to 0$ generates $\delta A^\alpha = c_{\beta\gamma}{}^{\alpha} A^\beta \Lambda^\gamma$ and since $\delta \mathscr{L} = 0$ the generalization of (4.2.4) tells us that

$$\Lambda^\alpha d(c_{\beta\alpha}{}^{\gamma} A^\beta \wedge {}^*F^\gamma + (b^\alpha)_{mn}\varphi^m \wedge {}^*(D\varphi)^n) = 0.$$

Thus the general conclusions which we drew from Maxwell's equations like the vanishing of the total charge in a closed universe or a plane wave still hold.

5. There is an essential difference to Maxwell's theory where the current was gauge invariant. Here the current does not transform like $\Omega$ under a gauge transformation, $J \not\to L^{-1}JL$ but transforms inhomogeneously. Indeed, $J \to L^{-1}JL$ would be incompatible with (4.2.7). Although ${}^*\Omega$ transforms like ${}^*\Omega \to L^{-1}{}^*\Omega L$, $d{}^*\Omega$ does not. Therefore it has no gauge invariant meaning to say that one has a certain amount of currents at a point $x$ in space, one can always find a gauge such that $J(x) = 0$. The total charges $Q_\alpha = \int_{N_3} {}^*J_\alpha = -\int_{\partial N_3} {}^*F_\alpha$ are more robust. Since they can be expressed as boundary integrals they transform homogeneously under all gauge transformations which have $dL = 0$ on $\partial N_3$.

The culprit responsible for the inhomogeneous transformation law of $J$ is the $A$ contained in the contribution from the gauge field. The contribution ${}^*J(\Phi) = b\Phi \wedge {}^*(D\Phi)$ from the scalar field transforms homogeneously. Indeed, if we put the current from the gauge field to the other side of (4.2.7) we obtain on the left-hand side the combination $d{}^*F^\alpha + c_{\beta\gamma}{}^{\alpha} A_\beta \wedge {}^*F^\gamma$ since the structure constants are totally antisymmetric, $c_{\beta\gamma}{}^{\alpha} = c_{\alpha\beta}{}^{\gamma}$. The extension (4.16.3) of the covariant derivative to $S_p(L(F))$ shows (compare Problem 7) that these are just the components of the covariant exterior derivative $D$ which is constructed such that it transforms with $L^{-1}D{}^*\Omega L$. Thus the Yang–Mills equation can also be written $D{}^*\Omega = -{}^*J(\Phi)$. ${}^*J(\Phi)$ alone is not conserved, however, one finds that $D$ applied twice to ${}^*\Omega$ gives zero (Problem 7) and thus $d{}^*J \neq 0$ but $D{}^*J(\Phi) = 0$. Together with (4.1.26) we find a form for (4.2.7) which looks like Maxwell's equation only that $d$ is replaced by $D$.

**Covariant Form of the Yang-Mills Equations** (4.2.9)

$$D\Omega = 0, \qquad D{*}\Omega = -{*}J(\Phi) \Rightarrow D{*}J(\Phi) = 0.$$

Since the gauge theories discussed so far seem to be the appropriate description for the electroweak and strong interactions, it is tempting to think that a theory of gravitation should follow the same pattern. According to the equivalence principle discussed in §1.3 gravity is related to the metric structure of space and time $M$. This suggests taking as fibre bundle $T(M)$, as fibre metric the pseudo-Riemannian metric, as gauge group the local Lorentz transformation and as gauge potential the connection $\omega$. Thus a possible Lagrangian would be $-\frac{1}{2}R^{\alpha}{}_{\beta} \wedge {*}R^{\beta}{}_{\alpha}$ + the matter Lagrangian (3.1.20). However, this does not give a theory of gravitation because the connection $\omega$ does not appear in (3.1.20), everything being expressible in terms of * and $d$. Thus the current ${*}J = \delta\mathscr{L}/\delta\omega$ does not get a contribution from the Maxwell or scalar fields. Only if we had a spinor field its spin density [21] would contribute to the current. However, one knows that all matter acts as source of the gravitational potential and so this theory does not seem to work. More specifically one believes that all energy and momentum currents are the source of gravitation. Since they are the generators of translations one might think that it is more the translation part of the Poincaré group and not the rotations which are relevant for gravity. If this is so we lose the strict analogy with the previous development where the gauge groups were orthogonal (or unitary) groups. Therefore the construction of the gravitational Lagrangian will involve some guess-work. In any case the energy and momentum currents of matter turn out to be $\partial\mathscr{L}^{\text{matter}}/\partial e^{\alpha}$. Here the tetrads $e^{\alpha} \in E_1$ are a basis and therefore it is reasonable to look at them as the gauge potentials replacing the $A$'s. In this case the $\omega$'s play the role of a field strength and not a potential, which they actually do in the equations (I: 1.1.6) for the motion of a particle in a gravitational field. If we use an orthogonal basis they contain all the information of the metric $g = \sum \eta_{\alpha\beta} e^{\alpha} \otimes e^{\beta}$ and its influence on matter comes only through the *-map contained in $\mathscr{L}^{\text{matter}}$. We shall represent matter by a scalar and the electromagnetic field and take the Lagrangian from (3.1.20) (in units $\rho e^2/m = 1$)

$$\mathscr{L}^{\text{matter}} = -\tfrac{1}{2}J \wedge {*}J - \tfrac{1}{2}F \wedge {*}F. \tag{4.2.10}$$

When choosing the gravitational Lagrangian $\mathscr{L}^{\text{gf}}$ some points should be born in mind:

**Remarks** (4.2.11)

1. At the time of the birth of gravitational theory, the requirement of general covariance provided some relief from the labor pains, but later on it was more often a source of confusion. The concept of a manifold incorporates

it automatically when the definition uses equivalence classes of atlases, and hence only chart-independent statements are regarded as meaningful. This program is by no means unique to gravitational theory—we have also followed it in classical mechanics and electrodynamics. The big difference is that now the metric $g$ on $M$ is not determined *a priori*.

2. There are some coordinate systems in which Einstein's equations simplify, just as Maxwell's equations are easier to work with in the Lorentz gauge $\delta A = 0$. For example, some formulas are shorter (cf. (4.2.20)) when written in the popular "harmonic coordinates," which satisfy $\delta\, dx = 0$ [26]. It is a matter of opinion, which we leave to the reader, whether this fact is of fundamental importance.
3. Even when the chart is fixed, there still remains the choice of a basis $e^i$. Orthogonal bases are special in that they standardize $g_{ik}$ as $\eta_{ik}$. They still leave open the possibility of a Lorentz transformation $e^\nu \to L^\nu{}_\beta(x)e^\beta$, $L^t(x)\eta L(x) = \eta$ for all $x$. $\mathscr{L}^{\text{matter}}$ in (4.2.10) is invariant under this transformation, which means that matter does not define a "teleparallelism." Looking just at matter one cannot observe the orientation of the local Lorentz systems relative to each other. It is thus reasonable to postulate that gravity does not prefer one frame either and therefore $\mathscr{L}^{\text{gf}}$ should be invariant under this transformation too.

We now have to select $\mathscr{L}^{\text{gf}} \in E_4$ from the material gathered in §4.1 for $T(M)$ and pose the requirements that it:

(a) is invariant under a change of basis; and
(b) is quadratic in the derivative of the $e^\alpha$.

This leads uniquely, up to a factor, to (for the last way of writing see Problem 8)

$$\mathscr{L}^{\text{g}} \simeq {}^*R = R_{\alpha\beta} \wedge {}^*e^{\alpha\beta} = 2d(e^\alpha \wedge {}^*de_\alpha) - (de^\alpha \wedge e^\beta) \wedge {}^*(de_\beta \wedge e_\alpha) + \tfrac{1}{2}(de^\alpha \wedge e_\alpha) \wedge {}^*(de^\beta \wedge e_\beta). \tag{4.2.12}$$

Thus the total Lagrangian will have three contributions. One from a scalar field $\mathscr{L}^{\text{s}} = -\frac{1}{2}J \wedge {}^*J$ which represents matter. For simplicity we choose units where $\rho e^2\, m = 1$. Secondly, the electromagnetic field adds $\mathscr{L}^{e} = -\frac{1}{2}F \wedge {}^*F$. Finally, comes $\mathscr{L}^{\text{g}}$ equipped with a factor to be adjusted later.

**Lagrangian for the Total System** (4.2.13)

$$\mathscr{L} = -\tfrac{1}{2}J \wedge {}^*J - \tfrac{1}{2}F \wedge {}^*F + \frac{1}{16\pi\kappa} R_{\alpha\beta} \wedge {}^*e^{\alpha\beta},$$

$$F = dA, \qquad J = dS + eA.$$

**Remarks** (4.2.14)

1. If we consider the $e^\alpha$ as counterpart to $A$ the simplest analog to $dA \wedge {}^*dA$ which is invariant under global Lorentz transformations would be $de^\alpha \wedge {}^*de_\alpha$. This leads to a theory with teleparallelisms and contradicts our information gained from experiments on the bending of light rays by the sun. The somewhat different form of (4.2.13) insures invariance under local Lorentz transformations. Because of this invariance we may use an orthonormal basis for the derivation of Euler's equation. The variation of the metric $g$ results from our making the variation of $e$ without imposing any orthogonality constraints. Since the metric $\eta$ is constant we may freely pull indices up and down under the derivative and use the notation $\omega_{\alpha\beta} = \eta_{\alpha\gamma}\omega^\gamma{}_\beta$ etc. It amounts only to a change of sign for each subscript zero.
2. Another quantity invariant under local gauge transformations would be simply $*1$. It contains no derivatives and by itself cannot lead to a differential equation. In an early stage of the theory Einstein added it in as a "cosmological term." Later it fell into disfavour and there is no empirical evidence for it. Yet its absence is still mysterious since there are many possible sources which could contribute such a term. We shall occasionally resurrect it for the purpose of comparison.
3. In an orthogonal basis we find that

$$\begin{aligned} {}^*e^{\alpha\beta} \wedge d\omega_{\alpha\beta} - d({}^*e^{\alpha\beta} \wedge \omega_{\alpha\beta}) &= -(d{}^*e^{\alpha\beta}) \wedge \omega_{\alpha\beta} \\ &= -\omega^\alpha{}_\gamma \wedge {}^*e^{\gamma\beta} \wedge \omega_{\alpha\beta} + \omega^\beta{}_\gamma \wedge {}^*e^{\alpha\gamma} \wedge \omega_{\alpha\beta} \\ &= -2{}^*e^{\beta\gamma} \wedge \omega_\gamma{}^\alpha \wedge \omega_{\alpha\beta}, \end{aligned}$$

and so we could simply use

$$\mathscr{L}' = \frac{-1}{16\pi\kappa} {}^*e^{\beta\gamma} \wedge \omega_\gamma{}^\alpha \wedge \omega_{\alpha\beta}$$

for gravitation. This equivalent $\mathscr{L}'$, when taken by itself, is altered by a change of basis. It is useful for showing why we get an equation of second order: it contains only $\omega$, i.e., derivatives of $e$, but not $d\omega$.

**Derivation of the Euler-Lagrange Equations** (4.2.15)

(a) Variation† of the Lagrangian of $S$: From $e^\alpha \wedge {}^*J = J \wedge {}^*e^\alpha$ (see (1.2.18(a))), we conclude that

$$e^\alpha \wedge \delta{}^*J = \delta J \wedge {}^*e^\alpha + J \wedge \delta{}^*e^\alpha - (\delta e^\alpha) \wedge {}^*J.$$

If we successively vary the $e$'s occurring in ${}^*e^\alpha$, we discover that

$$\delta{}^*e^\alpha = \delta e^\beta \wedge (i_\beta {}^*e^\alpha), \qquad i_\beta \equiv i_{e_\beta}.$$

† Here $\delta$ denotes the variation, not the codifferential.

Using this in the above formula, multiplying by $J_\alpha$, and summing:

$$J \wedge \delta^* J = \delta J \wedge {}^*J - \delta e^\alpha \wedge [J \wedge i_\alpha {}^*J + (i_\alpha J) \cdot {}^*J] \Rightarrow$$
$$\delta(-\tfrac{1}{2} J \wedge {}^*J) = -\delta J \wedge {}^*J + \tfrac{1}{2}\delta e^\alpha \wedge [J \wedge i_\alpha {}^*J + (i_\alpha J) \cdot {}^*J].$$

(b) Variation of the electromagnetic part of $\mathscr{L}$: As in part (a), we conclude from $e^{\alpha\beta} \wedge {}^*F = F \wedge {}^*e^{\alpha\beta}$ that

$$e^{\alpha\beta} \wedge \delta^* F = (\delta F) \wedge {}^*e^{\alpha\beta} + F \wedge \delta^* e^{\alpha\beta} - (\delta e^{\alpha\beta}) \wedge {}^*F.$$

As before, multiplication by $F_{\alpha\beta}$ yields

$$\delta(-\tfrac{1}{2} F \wedge {}^*F) = -\delta F \wedge {}^*F + \tfrac{1}{2}\delta e^\alpha \wedge [(i_\alpha F) \wedge {}^*F - F \wedge i_\alpha {}^*F].$$

(c) Variation of the gravitational part: First note that by (1.2.18(c)),

$$\delta^* e^{\alpha\beta} = \delta e^\gamma \wedge i_\gamma {}^* e^{\alpha\beta} = \delta e^\gamma \wedge {}^* e^{\alpha\beta}{}_\gamma, \qquad e^{\alpha\beta}{}_\gamma \equiv \eta_{\gamma\sigma} e^{\alpha\beta\sigma} \text{ etc.}$$

Although $R_{\alpha\beta}$ is itself constructed from the $\omega$'s and a variation of $e$ induces a variation of $\omega$, we need not calculate these variations, because

$$ {}^*e^{\alpha\beta} \wedge \delta R_{\alpha\beta} = d({}^*e^{\alpha\beta} \wedge \delta\omega_{\alpha\beta}) \tag{4.2.16}$$

(Problem 1), and hence the variation of $R$ does not affect the Euler-Lagrange equations.

The combination of these three results produces

$$\delta\mathscr{L} = -\delta J \wedge {}^*J - \delta F \wedge {}^*F + \delta e^\alpha$$
$$\wedge \left[ {}^*t_\alpha + {}^*\mathscr{T}_\alpha + \frac{1}{16\pi\kappa} {}^*e_{\alpha\beta\gamma} \wedge R^{\beta\gamma} \right] + \frac{1}{16\pi\kappa} d({}^*e^{\alpha\beta} \wedge \delta\omega_{\alpha\beta})$$
$${}^*t_\alpha = \tfrac{1}{2}((i_\alpha J){}^*J + J \wedge i_\alpha {}^*J),$$
$${}^*\mathscr{T}_\alpha = \tfrac{1}{2}((i_\alpha F) \wedge {}^*F - F \wedge i_\alpha {}^*F)$$

or, writing $J = dS + eA$ and $F = dA$,

$$\delta\mathscr{L} = \delta S\, d^*J - \delta A \wedge (d^*F + e^*J) + \delta e^\alpha$$
$$\wedge \left[ {}^*t_\alpha + {}^*\mathscr{T}_\alpha + \frac{1}{16\pi\kappa} {}^*e_{\alpha\beta\gamma} \wedge R^{\beta\gamma} \right] \tag{4.2.17}$$
$$- d\left( \delta S {}^*J + \delta A \wedge {}^*F - \frac{1}{16\pi\kappa} {}^*e^{\alpha\beta} \wedge \delta\omega_{\alpha\beta} \right).$$

The requirement that $\delta \int \mathscr{L} = 0$ therefore results in the

**Field Equations of the Total System** (4.2.18)

(a) $d^*J = 0$,
(b) $d^*F = -e^*J$,
(c) $-\tfrac{1}{2} {}^*e_{\alpha\beta\gamma} \wedge R^{\beta\gamma} = 8\pi\kappa({}^*t_\alpha + {}^*\mathscr{T}_\alpha)$.

**Remarks** (4.2.19)

1. These calculations generalize those of §3.1, because the metric has also been varied. The effect is to produce Einstein's equations (c) for the gravitational potential, in addition to the equations we got before.
2. The right sides of (c) are the energy-momentum currents of a scalar field and an electromagnetic field. Note that the gauge-dependent contribution (2.1.8; 1) of the canonical energy-momentum tensor does not appear, and that the Maxwellian contribution is coupled with gravity. These currents have the structure of the Hamiltonian $p\dot{q} - L$ in mechanics as anticipated in (1.3.22). They can be written

$$*t_\alpha = (i_\alpha *J) \wedge dS - i_\alpha \mathscr{L}^s, \qquad *\mathscr{T}_\alpha = i_\alpha *F \wedge dA - i_\alpha \mathscr{L}^e.$$

   Now $*J = -\partial \mathscr{L}^s / \partial\, dS$ (resp. $*F = -\partial \mathscr{L}^e / \partial\, dA$) correspond to $p$; $dS$ (resp. $dA$) correspond to $\dot{q}$ and the inner product picks out the right component.
3. Although we have used an orthogonal basis to derive (4.2.17), (c) has the same form in any basis, because of the homogeneous transformation law for $R_{\alpha\beta}$ (4.1.20; 3). Equation (c) seems to have a different structure than (a) and (b), it does not say that the derivative of a field strength is a current. (c) is the analog to the covariant form of the Yang–Mills equations (4.2.9). We shall shortly cast it into the form (4.2.7) which is like (a) and (b) but thereby the two sides of the equation lose the property of transforming homogeneously under gauge transformations.

**Different Versions of Einstein's Equations** (4.2.20)

(a) The classical version. The 3-forms of energy, momentum, and electric current all occur in (4.2.8). In §3.1 we wrote down the equations for the corresponding 1-forms, and we can similarly rewrite (c). By Rule (1.2.18(b)), for all $\omega \in E_2$,

$$-*e_{\alpha\beta\gamma} \wedge \omega = (*e_\alpha i_\beta i_\gamma + *e_\beta i_\gamma i_\alpha + *e_\gamma i_\alpha i_\beta)\omega.$$

   If $\omega = R^{\beta\gamma}$ and we observe that $i_\alpha i_\beta R^\beta{}_\gamma = i_\alpha R_\gamma = i_\gamma R_\alpha$ and $*e^\gamma i_\gamma R_\alpha = *R_\alpha$ (cf. (4.1.24; 3)), then we see that

$$R_\alpha - \tfrac{1}{2} e_\alpha R = 8\pi\kappa(t_\alpha + \mathscr{T}_\alpha),$$

   i.e.,

$$R_\alpha = 8\pi\kappa(t_\alpha + \mathscr{T}_\alpha - \tfrac{1}{2} e_\alpha(t + \mathscr{T})), \qquad t = i_\alpha t^\alpha, \text{ etc.}$$

   For the components of the Riemann-Christoffel tensor (4.1.24; 2) and the energy-momentum tensor $T_{\alpha\beta}$ such that $t_\alpha + \mathscr{T}_\alpha = T_{\alpha\beta} e^\beta$, this means that

$$R^\gamma{}_{\alpha\gamma\beta} - \tfrac{1}{2} g_{\alpha\beta} R^\gamma{}_{\sigma\gamma}{}^\sigma = 8\pi\kappa T_{\alpha\beta}.$$

(b) As a Yang Mills type equation. $\mathscr{L}'$ of (4.2.14; 3) can be written (see (4.2.12))

$$\mathscr{L}' = -\tfrac{1}{2}(de^\alpha \wedge e^\beta) \wedge {}^*(de_\beta \wedge e_\alpha) + \tfrac{1}{4}(de^\alpha \wedge e_\alpha) \wedge {}^*(de^\beta \wedge e_\beta)$$

if we momentarily put $8\pi\kappa = 1$. Performing the variational procedure (4.2.17) with the $e^\alpha$ we get another version of (4.2.18(c))

$$d{}^*F_\alpha = -{}^*J_\alpha$$

$${}^*F_\alpha = \frac{\partial \mathscr{L}}{\partial de^\alpha} = e^\beta \wedge {}^*(de_\beta \wedge e_\alpha) - \tfrac{1}{2}e_\alpha \wedge {}^*(de^\beta \wedge e_\beta) \tag{4.2.21}$$

$${}^*J_\alpha = \frac{\partial \mathscr{L}}{\partial e_\alpha} = i_\alpha {}^*J \wedge dS + i_\alpha {}^*F \wedge dA + i_\alpha {}^*F_\beta \wedge de^\beta - i_\alpha \mathscr{L}'.$$

They are now in the form of the exterior derivative of a field strength yielding the dual of a current and have the following features in common with the Yang–Mills equations:

(i) There is a contribution ${}^*\ell_\alpha = i_\alpha {}^*F_\beta \wedge de^\beta - i_\alpha \mathscr{L}'$ from the $e^\alpha$ to the currents. One can interpret it as the energy-momentum currents of the gravitational field. It has exactly the same structure $p\dot{q} - L$ as the contributions from $S$ and $A$.

(ii) Under $e^\alpha \to L^\alpha{}_\beta(x)e^\beta$ both $d{}^*F_\alpha$ and ${}^*J_\alpha$ transform inhomogeneously. From the three contributions to ${}^*J_\alpha$ it is only $\ell_\alpha$ which does not transform homogeneously. For any given point $x$ we can find a frame such that $\ell_\alpha(x) = 0$. Conversely, on a flat space with non-Cartesian coordinates $\ell^\alpha \neq 0$. As discussed in §1.3 these facts reflect how the balance of energy-momentum is affected in accelerated reference frames. The total currents are conserved, $d{}^*J_\alpha = 0$ and thus we draw again the conclusions that the total energy momentum of a closed universe or a periodic field is zero. If space is asymptotically flat and we restrict ourselves to transformations which are asymptotically constant Lorentz transformations then the total energy and momentum transform like a vector.

To see the relation to the classical version we have to re-express $de$ in terms of the $\omega$'s. This is done most easily if one remembers that the exterior product is the dual of the inner product

$$\begin{aligned} F^\alpha &= i_\beta(de^\beta \wedge e^\alpha - \tfrac{1}{2}\eta^{\alpha\beta}de^\gamma \wedge e_\gamma) \\ &= -i_\beta(\omega^\beta{}_\gamma \wedge e^{\gamma\alpha} - \tfrac{1}{2}\eta^{\alpha\beta}\omega^\gamma{}_\sigma \wedge e^\sigma{}_\gamma) \\ &= -\langle \omega_{\beta\gamma} | e^\beta \rangle e^{\gamma\alpha} + \tfrac{1}{2}\langle \omega_{\gamma\sigma} | e^\alpha \rangle e^{\sigma\gamma} \\ &= -\tfrac{1}{2} i_{\omega_{\beta\gamma}} e^{\alpha\beta\gamma} = -\tfrac{1}{2}{}^*(\omega_{\beta\gamma} \wedge {}^*e^{\alpha\beta\gamma}). \end{aligned} \tag{4.2.22}$$

Thus $d{}^*F^\alpha = \frac{1}{2}d(\omega_{\beta\gamma} \wedge {}^*e^{\alpha\beta\gamma})$ which has some similarity to the left side of (4.2.18(c)), the difference being ${}^*t^\alpha$. To get the latter in terms of the $\omega$'s we use ${}^*e^{\alpha\beta\gamma} = \varepsilon^{\alpha\beta\gamma\delta}e_\delta$,

$$d(\omega_{\beta\gamma} \wedge e_\delta) = d\omega_{\beta\gamma} \wedge e_\delta - \omega_{\beta\gamma} \wedge \omega_{\sigma\delta} \wedge e^\sigma$$

and

$$R_{\beta\gamma} = d\omega_{\beta\gamma} - \omega_{\sigma\beta} \wedge \omega^\sigma{}_\gamma.$$

Comparing (4.2.21) with (4.2.18(c)) we see

$${}^*t^\alpha = -\frac{\varepsilon^{\alpha\beta\gamma\delta}}{16\pi\kappa}(\omega_{\sigma\beta} \wedge \omega^\sigma{}_\gamma \wedge e_\delta - \omega_{\beta\gamma} \wedge \omega_{\sigma\delta} \wedge e^\sigma).$$

This form of the currents has been put forward by Landau and Lifshitz so that we shall call (4.2.21) the Landau–Lifshitz form of Einstein's equation. They did not use an orthogonal but a natural basis. In the natural basis $e^\alpha = dx^\alpha$, $\omega_{\alpha\beta} \equiv \Gamma_{\alpha\beta\mu}dx^\mu$ with $\Gamma_{\sigma\beta\mu} = \Gamma_{\sigma\mu\beta}$. Then the energy-momentum tensor of gravity, that is the components of $t^\alpha$ in this basis, are symmetric

$$dx^\rho \wedge {}^*t^\alpha = -\frac{\sqrt{|g|}}{16\pi\kappa}(\Gamma_{\sigma\beta\mu}\Gamma^\sigma{}_{\gamma\nu}\varepsilon^{\rho\mu\nu}{}_\delta + \Gamma_{\beta\gamma\mu}\Gamma_{\sigma\delta\nu}\varepsilon^{\rho\mu\nu\sigma}) \cdot \varepsilon^{\alpha\beta\gamma\delta} \cdot {}^*1$$

$$= dx^\alpha \wedge {}^*t^\rho.$$

This symmetry ensures local angular momentum conservation: $d({}^*t^\beta + {}^*\mathscr{T}^\beta + {}^*t^\beta) = 0$ implies $d(x^\alpha({}^*t^\beta + {}^*\mathscr{T}^\beta + {}^*t^\beta) - x^\beta({}^*t^\alpha + {}^*\mathscr{T}^\alpha + {}^*t^\alpha)) = 0$ only if it is true for $t$ as well as for $t$ and $\mathscr{T}$ (cf. (2.1.11; 2)) that $dx^\alpha \wedge {}^*t^\beta - dx^\beta \wedge {}^*t^\alpha = 0$. Here $x^\alpha$ is a local coordinate, so at this stage the theorem of conservation of angular momentum is formulated strictly on the domain of a single chart. The appropriate 3-form is defined globally only on special manifolds.

From (2.2.22) one deduces immediately (Problem 8) the so-called ADM-expression for the total energy in an asymptotically flat space. It was a major discovery [41] when it was shown that it is positive provided the external energies $\mathscr{T}^0$ and $t^0$ have this property. Thus the negative energy of gravitation can never exceed the energy of the sources if the space is to stay flat in the large.

**Remarks** (4.2.23)

1. The versions (4.2.20) do not exhaust the possibilities of writing Einstein's equations in terms of the exterior differentials of 2-forms. Numerous other variants have been proposed [10][29][39]. We just wanted to exhibit the analogy to non-abelian Yang–Mills theories. The field equations are nonlinear because the gravitational field carries energy and momentum. However these quantities evade localization since they transform inhomogeneously under local Lorentz transformations.

2. It is not the curvature forms $R_{\alpha\beta}$ but rather their contractions $R_\alpha$ that are locally determined by Einstein's equations. However, if the Weyl tensor is known, which happens, for instance, if the space is conformally flat ($C_{\alpha\beta} = 0$), then (4.2.11(c)) does determine $R_{\alpha\beta}$. In vacuo ($\mathcal{T}_\alpha = \ell_\alpha = 0$), $R_{\alpha\beta}$ and $C_{\alpha\beta}$ are the same and conformally flat solutions are flat. In two and three dimensions where the Weyl-tensor vanishes all solutions of Einstein's equations in vacuo are flat.
3. If one follows Cartan's suggestion and retains the torsion, generalizing the foregoing argument, then, like $R_\alpha$, it is determined locally, by the spin density of the matter present [21]. In the absence of spin, the space becomes torsionless, and the theory reduces to the one presented above. Since we know of no objects with sufficiently high spin densities, this variant of the theory agrees with experiment as well as Einstein's.
4. The geometrical significance of the 1-forms $R_\alpha$ determined by (4.2.20(a)) is brought out through the following heuristic argument: Suppose that $v$ points in the time direction $e^0$. Then if it is translated parallelly around the infinitesimal loop formed by $e^1 \wedge e^0$, it changes by $(\delta v)^1 = v^0 R^1{}_{010} \times$ the surface area, or, summing over the three spatial components,

$$(\delta v)^1 + (\delta v)^2 + (\delta v)^3 = v^0 R^\alpha{}_{0\alpha 0} \times \text{surface area}$$
$$= v^0 8\pi\kappa \int [T_{00} - \tfrac{1}{2} g_{00} T^\alpha{}_\alpha].$$

If $e^0$ is a geodesic vector field, then $v$ remains in the tangential direction of the geodesic vectors when translated parallelly along $e^0$. The positivity of the right side of the above equation indicates that the geodesic lines converge (Figure 55). This reflects the attractive character of gravity, and contains the seeds of the destruction of space and time that will be discussed in §4.6.

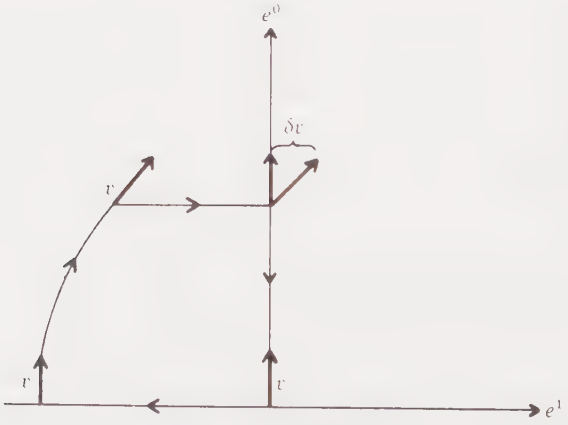



Figure 55 Parallel displacement of $v$ along the streamlines of $e^1$ and $e^0$

5. The $e^i$, and thus also the $\ell^\alpha$, can not in general be defined globally, though they can if there exists a compact, spatially orientable submanifold $N$ without a boundary. It is known that such a manifold is parallelizable, and hence $e^1$, $e^2$, and $e^3$ exist globally on $N$. If $e^0$ is taken as the timelike normal vector field, then $\ell^\alpha$ can be defined globally on $N$. Therefore the statement (1.3.35; 2) that an orientable, closed universe has zero total energy-momentum is always valid.

Now that the observables introduced in §1.3 have been identified, we are ready to derive the finite speed of propagation of the gravitational field, which, as anticipated in §1.1, equals the speed of light. Speaking mathematically, it is a matter of finding the characteristics of the equations (4.2.11(c)). In §1.3, the characteristics were defined as the possible

**Surfaces of Discontinuity of the Solutions** (4.2.24)

*Let $e^\alpha$ have possibly discontinuous second derivatives with respect to a local coordinate $u$, and let $t^\alpha$ and $\mathcal{T}^\alpha$ be continuous. Then either $R_{\alpha\beta}$ is continuous, or else $du \doteq n_\alpha e^\alpha$ in an orthogonal basis $e^\alpha$, where $n^2 \equiv n_\alpha n_\beta \eta^{\alpha\beta} = \langle du | du \rangle = 0$.*

**Remarks** (4.2.25)

1. If the $e^\alpha$ are allowed not to be $C^\infty$, then it is of course possible to choose them with discontinuous second derivatives even in flat space, a fact which reflects only the choice of basis. Any genuine discontinuity would have to show up in the $R_{\alpha\beta}$, and (4.2.13) states that such discontinuities can only occur along surfaces with lightlike normals.
2. The equations $\delta\, dA = J$ also allow arbitrary discontinuities in $A$ even if $J$ is continuous; after all, they can be contained in a gauge potential $A = d\Lambda$. The analogous alternatives are that either $F = dA$ is continuous, or else $n^2 = 0$ (cf. (2.2.1)).

**Proof of** (4.2.24)

The part of $de^\alpha$ that contains discontinuous first derivatives must be proportional to $du$:

$$de^\alpha = (A^\alpha{}_\beta + S^\alpha{}_\beta) du \wedge e^\beta,$$
$$A_{\alpha\beta} \equiv \eta_{\alpha\gamma} A^\gamma{}_\beta = -A_{\beta\alpha}, \qquad S_{\alpha\beta} = S_{\beta\alpha}.$$

We have separated the coefficients into a symmetric and an antisymmetric part, because they act differently in $\omega_{\alpha\beta} = -\omega_{\beta\alpha}$. If we accept the following equations modulo continuous terms, then

$$\omega_{\alpha\beta} = -A_{\alpha\beta}\, du + S_\alpha n_\beta - S_\beta n_\alpha, \qquad S_\alpha = S_{\alpha\gamma} e^\gamma,$$

as can be verified by substitution into $de^\alpha = -\omega^\alpha{}_\beta \wedge e^\beta$. By assumption, any possible discontinuity in the curvature originates with $d\omega_{\alpha\beta}$, hence either with $S'_\alpha$ such that $dS_\alpha = du \wedge S'_\alpha$ or with $dA_{\alpha\beta} = A'_{\alpha\beta}\, du$. The latter possibility does not contribute to $d\omega_{\alpha\beta}$, so the discontinuous part of the curvature becomes

$$R_{\alpha\beta} = du \wedge (S'_\alpha n_\beta - S'_\beta n_\alpha),$$

and thus

$$R_\beta \wedge du = i_\alpha R^\alpha{}_\beta \wedge du = (n^\alpha S'_\alpha n_\beta - S'_\beta n^2) \wedge du.$$

But in that case,

$$(R_\beta n_\gamma - R_\gamma n_\beta) \wedge du = n^2(S'_\gamma n_\beta - S'_\beta n_\gamma) \wedge du = n^2 R_{\gamma\beta}.$$

Since $R_\beta$ must be continuous according to Einstein's equations (4.2.11(c)), either $n^2 = 0$ or else $R_{\gamma\beta}$ stays continuous. □

**The Linear Approximation** (4.2.26)

The analogy we have just discovered between the characteristics of Einstein's and Maxwell's equations extends in the case of weak fields to a simple wave equation for the gravitational field. Even though we shall not show here when the contributions we drop are actually negligible, the approximation is of value as a first orientation to the problem, especially as the space around us is quite weakly curved.

Let

$$e^\alpha = dx^\alpha + \varphi^\alpha{}_\beta\, dx^\beta, \qquad \varphi_{\alpha\beta} \equiv \eta_{\alpha\sigma}\varphi^\sigma{}_\beta = \varphi_{\beta\alpha},$$

be an orthogonal basis, (the symmetry of $\varphi$ will be justified later) and suppose $|\varphi^\alpha{}_\beta(x)| \ll 1$ for all $x$. Then from

$$de^\alpha = \varphi^\alpha{}_{\beta,\gamma}\, dx^\gamma \wedge dx^\beta$$

it follows to first order in $\varphi$ that

$$\omega^\alpha{}_\beta = (\varphi^\alpha{}_{\gamma,\beta} - \varphi_{\beta\gamma,}{}^\alpha)dx^\gamma,$$

since (1.2.25) is satisfied to this order. In that case,

$$d\omega^\alpha{}_\beta = (\varphi^\alpha{}_{\gamma,\beta\rho} - \varphi_{\beta\gamma,}{}^\alpha{}_\rho)dx^\rho \wedge dx^\gamma,$$
$$i_\alpha\, d\omega^\alpha{}_\beta = (\varphi^\alpha{}_{\gamma,\beta\alpha} - \varphi_{\beta\gamma,}{}^\alpha{}_\alpha - \varphi^\alpha{}_{\alpha,\beta\gamma} + \varphi_{\beta\alpha,}{}^\alpha{}_\gamma)dx^\gamma.$$

In harmonic coordinates, where $\delta\, dx^\alpha = 0$,

$$\varphi_{\beta\alpha,}{}^\alpha = \tfrac{1}{2}\varphi^\alpha{}_{\alpha,\beta} \tag{4.2.27}$$

(Problem 2), and Einstein's equations (4.2.11(a)) become

$$-\varphi_{\beta\gamma,}{}^\alpha{}_\alpha\, dx^\gamma = 8\pi\kappa(T_{\beta\gamma} - \tfrac{1}{2}\eta_{\beta\gamma}T^\alpha{}_\alpha)dx^\gamma \tag{4.2.28}$$

in this approximation. These can be solved with the Green function (2.2.3):

$$\varphi_{\alpha\beta}(\bar{x}) = \varphi^{\text{in}}_{\alpha\beta}(\bar{x}) + 8\pi\kappa \int d^4x\, D^{\text{ret}}(\bar{x} - x)(T_{\alpha\beta}(x) - \tfrac{1}{2}\eta_{\alpha\beta}\, T^{\rho}{}_{\rho}(x))$$

$$\varphi^{\text{in}}_{\alpha\beta,\rho}{}^{\rho} = 0, \qquad \varphi^{\text{in}}_{\beta\alpha,}{}^{\alpha} = \tfrac{1}{2}\varphi^{\text{in}}{}^{\alpha}_{\alpha,\beta}. \tag{4.2.29}$$

**Remarks** (4.2.30)

1. The $T_{\alpha\beta}\, dx^{\beta}$ are the energy-momentum forms without the gravitational contribution. It is not inconsistent to neglect gravity in this approximation, because $\kappa T_{\beta,}{}^{\alpha}$ vanishes to zeroth order. This makes (4.2.29) and (4.2.27) consistent to first order.
2. In the static limit, $\int dt\, D^{\text{ret}}(x - \bar{x}) = \dfrac{1}{4\pi|\mathbf{x} - \bar{\mathbf{x}}|}$, $T_{\alpha\beta} = M j_{\alpha} j_{\beta}$, where $j = (-1, \mathbf{v})\delta^3(\mathbf{x})$, $\varphi_{\alpha\beta}$ becomes

$$\varphi_{\alpha\beta}(\bar{x}) = \frac{2M\kappa}{r}\,(j_{\alpha}j_{\beta} + \tfrac{1}{2}\eta_{\alpha\beta});$$

   i.e., the same result as stated in (I: 5.6.2), because $g = (\eta_{\alpha\beta} + 2\varphi_{\alpha\beta})dx^{\alpha} \otimes dx^{\beta}$. This justifies the choice of the factor $8\pi\kappa > 0$. The sign is not dictated by the geometry, but is only found empirically.
3. The analogy with electrodynamics should not make us overlook that, as discussed in volume I, chapter 6, the metric as measured is $g$ and not $\eta_{\alpha\beta}\, dx^{\alpha} \otimes dx^{\beta}$, although the difference is not great if the fields are weak.
4. The symmetry of $T_{\alpha\beta}$ justifies our ansatz $\varphi_{\alpha\beta} = \varphi_{\beta\alpha}$ a posteriori.

We close this section by investigating whether the generalization of the calculations of §2.1 and §3.1 connected with conservation laws brings new insights to the case of a gravitational field. Since the metric is not fixed a priori, there are now more invariance properties, and one would expect to find additional conserved quantities.

Returning to Equation (4.2.17) for the variation of $\mathscr{L}$, we start by looking at the new contribution from gravitation. If the variation comes from the Lie derivative $L_X$ in the direction of the vector field $X$, then Equation (4.2.17) implies that

$$\begin{aligned} & L_X(*e^{\alpha\beta} \wedge R_{\alpha\beta}) - d(*e^{\alpha\beta} \wedge L_X\omega_{\alpha\beta}) \\ &\quad = (L_X e_{\alpha}) \wedge *T^{\alpha} \\ &\quad = (i_X de_{\alpha} + d i_X e_{\alpha}) \wedge *T^{\alpha} \\ &\quad = -(i_X\omega_{\alpha\sigma})e^{\sigma} \wedge *T^{\alpha} + (i_X e^{\sigma})\omega_{\alpha\sigma} \wedge *T^{\alpha} - (i_X e_{\alpha})d*T^{\alpha} \\ &\qquad + d((i_X e_{\alpha})*T^{\alpha}), \end{aligned} \tag{4.2.31}$$

where $*T^{\alpha} \equiv *e^{\alpha\beta\gamma} \wedge R_{\beta\gamma} \in E_3$ and $e^{\alpha}$ is an orthogonal basis.

If (4.2.31) is integrated over a four-dimensional manifold $N$ without a boundary, and $X$ has compact support in $N$, then from $\int L_X\omega = 0$ for all

$\omega \in E_m$, the invariance of the integral under Lie differentiation (I: 2.6.11), and from $(i_X \omega_{\alpha\sigma}) e^\sigma \wedge {}^*T^\alpha = 0$ (because $e^\sigma \wedge {}^*T^\alpha = e^\alpha \wedge {}^*T^\sigma$) we infer that

$$\int_N x_\alpha (d{}^*T^\alpha + \omega^\alpha{}_\sigma \wedge {}^*T^\sigma) = 0, \qquad x_\alpha \equiv i_X e_\alpha .$$

Since this must be true for all vector fields $X$ of compact support in $N$ — no invariance properties have been assumed of $X$ — we obtain the

**Contracted Bianchi Identity** (4.2.32)

$$d{}^*T^\alpha = -\omega^\alpha{}_\sigma \wedge {}^*T^\sigma .$$

**Remarks** (4.2.33)

1. This fact follows from the more general equation (4.1.26) (Problem 3).
2. The invariance of the integral under general coordinate transformations is expressed by $\int L_X \omega = 0$. No new conservation theorems result from this general covariance, but only identities that hold independently of any field equation (cf. (2.1.8; 5) and (2.1.18; 3)).
3. Although (4.2.21) was derived for orthogonal bases, it has the same form in all bases, since the ${}^*T^\alpha$ transform as ${}^*e^\alpha$.
4. If Einstein's equations $T^\alpha = 16\pi\kappa(\mathcal{T}^\alpha + t^\alpha)$ hold, then (4.2.21) implies Equation (1.3.28):

$$d({}^*\mathcal{T}^\alpha + {}^*t^\alpha) = -\omega^\alpha{}_\beta \wedge ({}^*\mathcal{T}^\beta + {}^*t^\beta).$$

5. Because of the symmetry $i_\beta T_\alpha = i_\alpha T_\beta$, the 1-forms $T_\alpha$ have 10 linearly independent components, among which (4.2.21) creates 4 differential identities. Thus only 6 of Einstein's equations are independent of one another. This is felt to be the correct number of equations for the 10 components $g_{\alpha\beta} = g_{\beta\alpha}$ of $g = g_{\alpha\beta}\, dx^\alpha \otimes dx^\beta$: the equations ought not to fix the coordinate system, and so there must remain 4 arbitrary functions $\bar{x}^\alpha(x)$ to play with.
6. It is part of the relativity folklore that Einstein's theory differs from other field theories inasmuch as Einstein's equations also determine the equations of motion of matter. In particular (4.2.32) is supposed to imply that particles move on geodesics provided no other forces act on them. Whereas the proof of this claim for point particles encounters the difficulty that they generate singular gravitational fields we shall find a simple proof for continuous matter in (4.6.30; 4). There we will show that for an ideal fluid without pressure Bianchi's identity demands that the velocity field be geodesic. Similarly, if one has only one scalar field (4.2.32) implies the field equation. With two scalar fields the condition (4.2.32) for the

sum of their energy momentum tensors can not imply the field equation because they may interact without changing the conservation of their total energy-momentum. Thus the situation is not too different from Maxwell's equations where for one real field describing charged matter $d^*J = 0$ is already the field equation $d^*(dS + eA) = 0$. For one complex field $\sqrt{\rho}\, e^{iS}$ the current is $\rho(dS + eA)$ and its conservation does not imply the field equation.

**Problems** (4.2.34)

1. Show that $^*e^{\rho\tau} \wedge \delta R_{\rho\tau} = d(^*e^{\rho\tau} \wedge \delta\omega_{\rho\tau})$.
2. Show that $\delta\, dx^\alpha = 0$ implies that $\varphi_{\alpha\beta,}{}^{\beta} = \frac{1}{2}\varphi^{\beta}{}_{\beta,\alpha}$ for weak fields.
3. Derive (4.2.32) from (4.1.26).
4. What effect would a term $\Lambda^*\mathbf{1}$ in $\mathscr{L}$ have on Einstein's equations?
5. Calculate $^*\ell_0$ from (4.2.18(c)) on flat space in the orthogonal basis of polar coordinates.
6. Show that $R^\alpha{}_\beta \wedge R^\beta{}_\alpha$ is exact.
7. Use the natural extension of $D$ for $V \in \mathrm{Sp}(L(F))$: $DV = dV + \omega \wedge V + (-)^p V \wedge \omega$ to show $DD^*\Omega = 0$.
8. Show

$$\begin{aligned}\tfrac{1}{2}R_{\alpha\beta} \wedge {}^*e^{\alpha\beta} = {} & d(e^\alpha \wedge {}^*de_\alpha) - \tfrac{1}{2}(de^\alpha \wedge e^\beta) \wedge {}^*(de_\beta \wedge e_\alpha) \\ & + \tfrac{1}{4}(de^\alpha \wedge e_\alpha) \wedge {}^*(de^\beta \wedge e_\beta).\end{aligned}$$

   *Hint*: Use (4.2.22) and (4.2.14; 3).

9. Show that the total energy ("ADM-energy")

$$\frac{-1}{8\pi\kappa}\int_N d^*F^0 = \frac{-1}{8\pi\kappa}\int_{\partial N} {}^*F^0,$$

   where $\partial N = \{x \in \mathbb{R}^4 : |\vec{\mathbf{x}}| = R,\ x_0 = 0\}$ is an asymptotically flat space is given by

$$\lim_{R\to\infty} \frac{1}{16\pi\kappa}\int_{\partial N} {}^*e^{j\rho}(g_{jk,k} - g_{kk,j}) \qquad \text{with} \quad j, k = 1, \ldots, 3.$$

   (Use a natural basis $e^k = dx^k$, (4.1.37) for $\omega$ and that $g_{ik} \to \eta_{ik}$ for $R \to \infty$.)

**Solutions** (4.2.35)

1. $$\begin{aligned} d(^*e^{\rho\tau} \wedge \delta\omega_{\rho\tau}) &= -2\omega^\rho{}_\sigma \wedge {}^*e^{\sigma\tau} \wedge \delta\omega_{\rho\tau} + {}^*e^{\rho\tau} \wedge d\delta\omega_{\rho\tau} \\ &= {}^*e^{\rho\tau} \wedge (d\delta\omega_{\rho\tau} + 2\omega_\rho{}^\sigma \wedge \delta\omega_{\sigma\tau}) = {}^*e^{\rho\tau} \wedge \delta R_{\rho\tau}. \end{aligned}$$

2. To first order, $dx^\alpha = e^\alpha - \varphi^\alpha{}_\beta e^\beta$; so

$$0 = {}^*\delta\, dx^\alpha = d({}^*e^\alpha - \varphi^\alpha{}_\beta {}^*e^\beta) = -\omega^\alpha{}_\beta \wedge {}^*e^\beta - \varphi^\alpha{}_{\beta,\gamma} e^\gamma \wedge {}^*e^\beta$$
$$= (\varphi_{\beta\gamma}{}^\alpha - \varphi^\alpha{}_{\gamma,\beta} - \varphi^\alpha{}_{\beta,\gamma}) e^\gamma \wedge {}^*e^\beta.$$

By (1.2.28), however, $e^\gamma \wedge {}^*e^\beta = \eta^{\gamma\beta} {}^*\mathbf{1}$.

3. $$d{}^*T^i = d({}^*e^{imk} \wedge R_{mk}) = -\omega^i{}_j \wedge {}^*e^{jmk} \wedge R_{mk} - 2\omega^m{}_s \wedge {}^*e^{iskh} \wedge R_{mk}$$
$$+ 2{}^*e^{imk} \wedge \omega_m{}^s \wedge R_{sk} = -\omega^i{}_j \wedge {}^*T^j.$$

4. $\delta(\Lambda {}^*\mathbf{1}) = \Lambda \delta e^j \wedge i_j {}^*\mathbf{1} = \Lambda \delta e^j \wedge {}^*e_j$ (see (1.2.18(c)). Thus there is an additional term $\sim {}^*e_j$ in ${}^*\mathscr{T}_j$.

5. $e^\alpha = (dt, dr, r\, d\vartheta, r \sin\vartheta\, d\varphi)$, $\omega_{12} = -d\vartheta$, $\omega_{23} = -\cos\vartheta\, d\varphi$, $\omega_{31} = \sin\vartheta\, d\varphi$,
$\frac{1}{2}\varepsilon_{0\beta\gamma\delta} d(\omega_{\beta\gamma} \wedge e_\delta) = d(-\cos\vartheta\, d\varphi \wedge dr + 2r \sin\vartheta\, d\varphi \wedge d\vartheta)$
$= -\sin\vartheta\, dr \wedge d\vartheta \wedge d\varphi.$

6. $R^\alpha{}_\beta \wedge R^\beta{}_\alpha = d[\omega^\alpha{}_\gamma \wedge d\omega^\gamma{}_\beta + \frac{2}{3}\omega^\alpha{}_\beta \wedge \omega^\beta{}_\gamma \wedge \omega^\gamma{}_\alpha]$.

7. $D{}^*\Omega = d{}^*\Omega + \omega \wedge {}^*\Omega - {}^*\Omega \wedge \omega$
$DD{}^*\Omega = d\omega \wedge {}^*\Omega - {}^*\Omega \wedge d\omega - \omega \wedge d{}^*\Omega - d{}^*\Omega \wedge \omega$
$+ \omega \wedge (d{}^*\Omega + \omega \wedge {}^*\Omega - {}^*\Omega \wedge \omega) + (d{}^*\Omega + \omega \wedge {}^*\Omega - {}^*\Omega \wedge \omega) \wedge \omega$
$= \Omega \wedge {}^*\Omega - {}^*\Omega \wedge \Omega = 0$

according to (1.2.17).

8. According to (4.2.14; 3) ${}^*e^{\alpha\beta} \wedge R_{\alpha\beta} = d({}^*e^{\alpha\beta} \wedge \omega_{\alpha\beta}) + {}^*e^{\beta\gamma}\omega_\gamma{}^\alpha \wedge \omega_{\alpha\beta}$. Now always using (1.2.17),

$$e_\alpha \wedge {}^*de^\alpha = {}^*e^\beta \langle \omega^\alpha{}_\beta | e_\alpha \rangle$$

and

$$({}^*e^{\alpha\beta}) \wedge \omega_{\alpha\beta} = {}^*i_{\omega_{\alpha\beta}} e^{\alpha\beta} = 2{}^*e^\beta \langle \omega^\alpha{}_\beta | e_\alpha \rangle \Rightarrow$$
$$\tfrac{1}{2}{}^*e^{\alpha\beta} \wedge \omega_{\alpha\beta} = e^\alpha \wedge {}^*de^\alpha.$$

Next with (4.2.22)

$$-de^\alpha \wedge {}^*F_\alpha = \tfrac{1}{2}\omega_{\alpha\sigma} \wedge e^\sigma \wedge \omega_{\beta\gamma} \wedge {}^*e^{\alpha\beta\gamma}$$
$$= \tfrac{1}{2}\omega_{\alpha\sigma} \wedge \omega_{\beta\gamma}[\eta^{\sigma\alpha} {}^*e^{\beta\gamma} - \eta^{\sigma\beta} {}^*e^{\alpha\gamma} + \eta^{\sigma\gamma} {}^*e^{\alpha\beta}]$$
$$= {}^*e^\alpha \wedge \omega_\gamma{}^\beta \wedge \omega_{\alpha\beta}.$$

Inserting the expression (4.2.21) for ${}^*F_\alpha$ completes the demonstration of the equality.

9. With (4.2.22) we have

$$-{}^*F^0 = -\tfrac{1}{2}(g_{ik,j} + g_{ij,k} - g_{jk,i})\tfrac{1}{2}e^j \wedge {}^*e^{ik0} = \tfrac{1}{2}(-g_{ii,k} + g_{ik,i}) {}^*e^{0k}$$

since by symmetry the first term cancels out and the two others give the same contribution.

## 4.3 Maximally Symmetric Spaces

*The spaces with the simplest structure, after flat spaces, are those of constant curvature. They are a generalization of the spherical surface and, though simple, have some physically interesting aspects.*

Killing vector fields generate isometries (i.e., diffeomorphisms that leave $g$ invariant) of the space and are bijectively related to the constants of motion and conserved currents. Yet the fields need not be complete; it is possible for their flow to lead out of the manifold (see (I: 2.3.7)). However, they imprint a local structure on the space even when they do not generate one-parameter groups of isometries. Generally there are none but if there are enough of them around, Einstein's equations become more tractable, and explicit calculations are possible.

The prototype of a Killing vector field is a rotation of $\mathbb{R}^m$; for some pair of indices $(i, k)$, $v^i = x^k$ and $v^k = -x^i$, and the other components are zero. Note that

(a) $v^i{}_{,k} + v^k{}_{,i} = 0$, and
(b) $v^l{}_{,jk} = 0$.

The generalizations of these facts to pseudo-Riemannian spaces are

**The Relationships among the Covariant Derivatives of Killing Vector Fields** $v$ (4.3.1)

*We use the natural basis* $e_\alpha = \partial_\alpha$ *and the notation*

$$\langle D_{e_\beta} v | e_\alpha \rangle \equiv v_{\alpha;\beta}, \qquad \langle D_{e_\gamma} D_{e_\beta} v | e_\alpha \rangle \equiv v_{\alpha;\beta;\gamma}.$$

*Then, with the* $R^\lambda{}_{\sigma\rho\mu}$ *of* (4.1.24; 2),

(a) $v_{\alpha;\beta} + v_{\beta;\alpha} = 0$, *and*
(b) $v_{\mu;\rho;\sigma} = R^\lambda{}_{\sigma\mu\rho} v_\lambda$.

**Proof**

(a) Killing vector fields leave the scalar product invariant, $L_v\langle X | Y \rangle = \langle L_v X | Y \rangle + \langle X | L_v Y \rangle$. On the other hand, with Axioms (4.1.6(f)) and (g′),

$$\begin{aligned} L_v\langle X | Y \rangle &= D_v\langle X | Y \rangle = \langle D_v X | Y \rangle + \langle X | D_v Y \rangle \\ &= \langle D_X v | Y \rangle + \langle X | D_Y v \rangle + \langle L_v X | Y \rangle + \langle X | L_v Y \rangle. \end{aligned}$$

Combined, these make $\langle D_X v | Y \rangle + \langle X | D_Y v \rangle = 0$; i.e., (a), if $X$ and $Y$ are taken as the basis fields.

(b) Since a Killing vector field $v$ conserves the metric structure and a torsion-free connection is uniquely determined by the metric $L_v$ commutes with the covariant exterior derivative $D$. To demonstrate this formally one has to consider the one parameter (local) group $\Phi_t$ of isometries generated by $v$ and consider

$$\frac{\partial}{\partial t} D\Phi_t \Big|_{t=0} = \frac{\partial}{\partial t} \Phi_t D \Big|_{t=0}.$$

Thus we infer from (4.1.39) that $R(X, v)u = \langle D_X Dv | u \rangle$. Using the properties (4.1.44; 2) of $R$ this becomes (b) when written in index notation.

**Discussion** (4.3.2)

(b) allows the second derivative of $v$ to be written linearly in $v$. By carrying the procedure further, one can reduce all the higher derivatives to $v$ and its first derivatives. If we assume analyticity, we can express $v$ locally in terms of $v$ and its first derivative at a single point, say 0:

$$v_\rho(x) = A_\rho{}^\lambda(x) v_\lambda(0) + B_\rho{}^{\lambda\sigma}(x) v_{\lambda;\sigma}(0).$$

This greatly restricts the number of possible Killing vector fields. In an $m$-dimensional space, $v_{\lambda;\sigma}(0) = -v_{\sigma;\lambda}(0)$ can assume $m(m-1)/2$ values, and $v_\lambda(0)$ can assume $m$ values. Hence there are at most $m + m(m-1)/2 = m(m+1)/2$ independent Killing vector fields.

**Remarks** (4.3.3)

1. "Independent" means that they have no linear relationships with constant coefficients. They may satisfy equations with variable coefficients as rotations $x_j \partial_i - x_i \partial_j$ are expressed by translations $\partial_i$. Since the Killing vector fields do not form a module (2.1.10; 1), this does not mean dependence.
2. On flat space the Euclidean group is the largest group of isometries (I: 4.1.13; 4), and it has exactly $m(m+1)/2$ parameters. The statement that the group can be at most this large if space is curved is therefore quite plausible.

To classify the symmetric spaces, we begin with a

**Definition** (4.3.4)

(a) A space is **maximally symmetric** iff it possesses $m(m+1)/2$ independent Killing vector fields.

(b) A space is **isotropic about the point** $x$ iff it has $m(m-1)/2$ Killing vector fields, for which $x$ is a fixed point of the flow, and the $A^i{}_k$ of $(L_v e^i)(x) = A^i{}_k(x)e^k(x)$, where $e^k$ are an orthogonal basis, generate the total Lorentz group of $(\mathbb{R}^m, \eta)$.
(c) A space is **isotropic** iff it is isotropic about all of its points.
(d) A space is **homogeneous** if it has a transitive† group of isometries.
(e) A space is **stationary** if it has a timelike Killing vector field. If the latter is orthogonal to a family of spacelike hypersurfaces the space is called **static**.

**Remarks** (4.3.5)

1. Definition (b) means that there are $m(m-1)/2$ Killing vector fields $v^i$ such that $v^i{}_x(x) = 0$, and the $v^i{}_{x;\beta}(x)$ form a basis for the space of $m \times m$ antisymmetric matrices, where $i$ runs from 1 to $m(m-1)/2$.
2. On a maximally symmetric space there are Killing vector fields as in (b) at every point. Therefore it is isotropic.
3. To appreciate the distinction made in (e) we remark that in general for a vector field $X$ there is locally a family of transversal hypersurfaces $N_x$ such that $X$ and the vectors in $N_x$ span the tangent space. However, this hypersurfaces will not be orthogonal to $X$ in the sense of a given metric and there will be no orthogonal family of hypersurfaces. In contradistinction to the above for a given spacelike hypersurface one can find, at least locally, an orthogonal geodesic vector field.

The curvature must be the same in every direction on an isotropic space. Because of this, there is an extremely simple

**Structure of the Curvature Forms of an Isotropic Space** (4.3.6)

*On an isotropic space,*

$$R^{ik} = Ke^{ik}, \textit{ for some constant } K.$$

**Proof**

Let $\Phi$ be an isometry and $\Phi_*$ the translation it induces on tensors. We noted that $\Phi_*$ commutes with $D$ and from $DD\Phi_* = \Phi_* DD$ we conclude $R\Phi_* = \Phi_* R$. Let $\Phi$ be an isometry with $x$ as fixed point such that $(\Phi_* e^i)(x) = L^i{}_k e^k(x)$ with $L^t \eta L = \eta$. If $R$ is decomposed in this basis then its invariance under $\Phi_*$ means that the Riemann-Christoffel-tensor (4.1.42; 2)

† Transitive means that any point can be reached from every point by a transformation from the group.

is invariant under $R_{ijkl} \to R_{mnop}L^m{}_i L^n{}_j L^o{}_k L^p{}_l$, it transforms as a matrix in the antisymmetric tensor product space. This representation is irreducible (except for $m = 4$) and the matrix group elements must be proportional to the unit matrix: $R_{ij}{}^{lm} = K(\delta_i{}^l\delta_j{}^m - \delta_i{}^m\delta_j{}^l)$ or for the curvature forms $R_{ij}(x) = K(x)e_{ij}(x)$. For $m = 4$ there would be the possibility $R_{ij} = K^*e_{ij}$ which is excluded by $R_{ij} \wedge e^j = 0$. In order to see why $K$ has to be constant, consider the equation $dR_{ik} = dK \wedge e_{ik} + K\, de_{ik}$. By using Bianchi's identity (4.1.26), we discover that $dK = 0$, and hence $K$ is independent of $x$. (For $m = 2$ another argument works.) □

**Remarks** (4.3.7)

1. If the curvature is independent of the direction, then it is also independent of the position. For that reason, one says simply that such spaces have constant curvature.
2. By (4.1.44; 3) and (4.3.6), isotropic spaces have vanishing Weyl forms, and hence they are conformally flat if $m > 3$.
3. It is even true that the existence of more than $m(m-1)/2 + 2$ Killing vector fields implies that $R_{ik} = Ke_{ik}$ (cf. [4]).

**Construction of Isotropic Spaces with** $m \geq 3$ (4.3.8)

Since such spaces are conformally flat, there is an orthogonal basis of the form $e^a = dx^a/\psi$, $\psi \in E_0$. Therefore $de^a = \psi_{,b}e^{ba}$, and thus $\omega_{ab} = \psi_{,a}e_b - \psi_{,b}e_a$. With the help of (4.1.45) again, this leads to the curvature forms

$$R_{ab} = \psi(\psi_{,ac}e^c{}_b - \psi_{,bc}e^c{}_a) - \psi_{,c}\psi^{,c}e_{ab}.$$

If this is to equal $Ke_{ab}$, then $\psi_{,ab}$ must be zero for all $a \neq b$, and hence

$$\psi = \sum_{i=1}^{m} f^i(x^i).$$

If $f_a = \eta_{ai} f^i$, then (4.3.6) implies that

$$f_b'' + f_a'' = \psi^{-1}(K + f_c' f^{c\prime}).$$

Since the left side depends only on $x^a$ and $x^b$, while the right side is the same for all $a$ and $b$, both sides are in fact constant. This makes $f$ a quadratic function, and therefore $\psi$ can be put into the form

$$\psi = 1 + \frac{K}{4} x^a x^b \eta_{ab}.$$

Therefore, locally a space of constant curvature always has coordinates for which

$$g = \frac{dx^i\, dx^k\, \eta_{ik}}{(1 + Kx^2/4)^2}, \qquad e^i = \frac{dx^i}{1 + Kx^2/4},$$
$$\omega^{ik} = \frac{K}{2}(x^i e^k - x^k e^i) \tag{4.3.9}$$

(where $x^2 \equiv x^a x^b \eta_{ab}$).

**The Killing Vector Fields of Isotropic Spaces** (4.3.10)

Because of the isotropy about the origin, the generators of rotations are Killing vector fields: Let $v = x_j \partial_k - x_k \partial_j$, where $x_j = \eta_{jk} x^k$ and $\partial_k$ is the dual basis to $dx^k(i_{\partial_k} dx^i = \delta^i{}_k)$, and fix a pair of indices $(j, k)$; then

$$\begin{aligned} L_v e^m &= d i_v e^m + i_v de^m \\ &= d\left(x_j \left(1 + \frac{Kx^2}{4}\right)^{-1} i_{\partial_k} dx^m\right) \\ &\quad - x_j \frac{K}{2}\left(1 + \frac{Kx^2}{4}\right)^{-2} x_l i_{\partial_k} dx^l \wedge dx^m - (j \leftrightarrow k) \\ &= d\left[x_j \delta^m{}_k \left(1 + \frac{Kx^2}{4}\right)^{-1}\right] \\ &\quad + \frac{K}{2} x_j \left(1 + \frac{Kx^2}{4}\right)^{-2} (\delta^m{}_k x_l\, dx^l - x_k\, dx^m) - (j \leftrightarrow k) \\ &= dx_r \left(1 + \frac{Kx^2}{4}\right)^{-1} (\delta^m{}_k \delta^r{}_j - \delta^m{}_j \delta^r{}_k). \end{aligned}$$

Since $L_v e^m = A^{mr} e_r$ and $A^{mr} = \delta^m{}_k \delta^r{}_j - \delta^m{}_j \delta^r{}_k = -A^{rm}$, rotations about the origin are Killing vector fields. (See (2.1.10; 2).)

Moreover, there are generalizations of translations: If $k$ is fixed, then $v = \partial^k(1 - x^2 K/4) + (K/2)x^k x_j \partial^j$ is a Killing vector field. This can be verified as in (4.3.10), or, what is easier in this case, one can verify the equation of Problem 1 for the components $v^i$ and $g_{lm}$ of $v$ and $g$ in the natural basis,

$$\begin{aligned} 0 &\stackrel{?}{=} v^i g_{lm,i} + g_{im} v^i{}_{,l} + g_{il} v^i{}_{,m} \\ &= \left(1 + \frac{Kx^2}{4}\right)^{-3} \left\{ -\eta_{lm}\left[Kx^k\left(1 - \frac{x^2 K}{4}\right) + \frac{x^k x^2 K^2}{2}\right] \right. \\ &\quad + \left(1 + \frac{Kx^2}{4}\right)\left[-\frac{K}{2}\delta^k{}_m x_l + \frac{K}{2}(x^k \eta_{ml} + x_m \delta^k{}_l) - \frac{K}{2}\delta^k{}_l x_m \right. \\ &\quad \left. \left. + \frac{K}{2}(x^k \eta_{ml} + x_l \delta^k{}_m)\right]\right\} \\ &= 0. \end{aligned}$$

These $v$'s form a basis for small $x$, and so the group they generate in a neighborhood of the origin acts transitively: any point can be sent to any other. We can assemble our discoveries in a

**Proposition** (4.3.11)

*For a pseudo-Riemannian manifold $M$, the following properties are equivalent:*

(a) *$M$ is maximally symmetric;*
(b) *$M$ has constant curvature;*
(c) *$M$ is isotropic;*
(d) *$M$ is homogeneous and isotropic about some point.*

**Remark** (4.3.12)

Our arguments have been strictly local, so nothing can be concluded about the global behavior—The Killing vector fields need not even be complete. Unions and pieces of spherical surfaces are also isotropic spaces. However, isotropic spaces with the same $K$ are locally isometric.

Every $m$-dimensional manifold can be imbedded as a submanifold in $\mathbb{R}^{2m+1}$. Isotropic spaces can be imbedded in $\mathbb{R}^{m+1}$, whereby the metric is the restriction of a pseudo-Euclidean metric on $\mathbb{R}^{m-1}$. (Restricting the metric means that the scalar product determining it is restricted to the vectors in the tangent space of the submanifold.)

**The Geometrical Imbedding of Isotropic Spaces** (4.3.13)

*Let $\eta_{\alpha\beta}$, $\alpha, \beta = 0, 1, \ldots, m-1$, fix the sign of the metric of an isotropic space, and choose the curvature $|K|$ as the unit of length, so that $K = \pm 1$. Let the $\eta_{ik}$ on $\mathbb{R}^{m+1}$ equal $\eta_{\alpha\beta}$ for $i, k = 0, 1, \ldots, m-1$, and let $\eta_{mm} = K$. Then the isotropic space is locally isometric to the submanifold $H = \{\bar{x} \in \mathbb{R}^{m+1} : \bar{x}^i \bar{x}^k \eta_{ik} = K\}$, where $g$ is the restriction of $d\bar{x}^i \otimes d\bar{x}^k \eta_{ik}$ to $H$.*

**Proof**

The equation for $H$ can be written as ($m = \dim H$)

$$K\bar{x}^m\bar{x}^m = K - \eta_{\alpha\beta}\bar{x}^\alpha\bar{x}^\beta \equiv K - \bar{r}^2.$$

Introduce the coordinates $x^\alpha \in \mathbb{R}^m$ on $H$ by

$$\bar{x}^\alpha = \frac{x^\alpha}{1 + Kr^2/4}, \qquad \bar{x}^m = \frac{1 - Kr^2/4}{1 + Kr^2/4},$$

where $r^2 = x^\alpha x^\beta \eta_{\alpha\beta}$; then

$$d\bar{x}^\alpha = \frac{dx^\alpha(1 + Kr^2/4) - Kx^\alpha r\, dr/2}{(1 + Kr^2/4)^2},$$

$$d\bar{x}^m = \frac{-Kr\, dr}{(1 + Kr^2/4)^2},$$

and the restriction of the metric

$$d\bar{x}^i\, d\bar{x}^k\, \eta_{ik} = d\bar{x}^\alpha\, d\bar{x}^\beta\, \eta_{\alpha\beta} + K\, d\bar{x}^m\, d\bar{x}^m = dx^\alpha\, dx^\beta\, \eta_{\alpha\beta}\left(1 + \frac{Kr^2}{4}\right)^{-2}$$

takes on the form (4.3.9). □

**Remarks** (4.3.14)

1. The Killing vector fields of (4.3.10) are simply the restrictions of the generators $\bar{x}^k\, \bar{\partial}^j - \bar{x}^j\, \bar{\partial}^k$ of the Lorentz group of $\mathbb{R}^{m+1}$ to $H$. Their flow leaves $H$ and the metric $\eta_{ik}\, d\bar{x}^i\, d\bar{x}^k$ invariant, and hence so does their restriction to $H$. The group of isometries of spaces of maximal symmetry is therefore isomorphic to some Lorentz group $O(n, m + 1 - n)$.
2. The geodesics are the intersections of $H$ with hyperplanes passing through the origin. They are the solutions of a variational problem $\delta \int ds\, \dot{\bar{x}}^i \dot{\bar{x}}^k \eta_{ik} = 0$ with the constraint $\bar{x}^i \bar{x}^k \eta_{ik} = K$. The introduction of a Lagrange multiplier $\lambda$ leads to the equations $\ddot{\bar{x}}^i = \lambda \bar{x}^i$, making the

$$L^{ik} = \bar{x}^i \dot{\bar{x}}^k - \bar{x}^k \dot{\bar{x}}^i$$

constant. Thus $\bar{x}$ lies on the plane of $\bar{x}(0)$, $\dot{\bar{x}}(0)$ since

$$0 = \varepsilon_{ike} \bar{x}^i L^{ke} = \varepsilon_{ihe} \bar{x}^i \dot{\bar{x}}^k x^l.$$

**The Physical Significance of Isotropic Spaces** (4.3.15)

Spaces of a high degree of symmetry gratify the esthetic feelings of physicists, and are therefore popular as models of the world. Since the space around us is isotropic and homogeneous as far as we can see, its isotropy about every point is often elevated to a cosmological principle. Some theorists went beyond this and require a maximal symmetry for space and time, the "perfect cosmological principle." But aside from such cosmological speculation, maximally symmetric spaces are also important as solutions of Einstein's equations when the energy-momentum distribution is sufficiently symmetric. It is necessary to distinguish between the cases of positive and negative $K$ ($K = 0$ is Minkowski space):

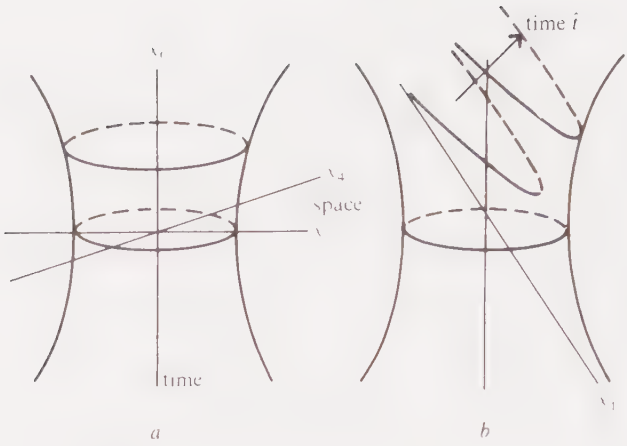



Figure 56 Various cross-sections of the de Sitter universe

$K = 1$ **The de Sitter Universe** (4.3.16)

This can be represented as the hyperboloid

$$-x_0^2 + x_1^2 + x_2^2 + x_3^2 + x_4^2 = 1$$

in $\mathbb{R}^5$ with $g = -dx_0^2 + dx_1^2 + dx_2^2 + dx_3^2 + dx_4^2$. When reduced to $\mathbb{R}^3$, it looks as in Figure 56. The intersections with planes containing the $x_0$-axis are timelike geodesics. If these geodesics are introduced as coordinate lines with the proper time as a new coordinate $t$ (these are known as **synchronous**, or **comoving**, coordinates; see Problem 5),

$$x_0 = \sinh t, \qquad x_1 = \cosh t \sin \chi \sin \vartheta \cos \varphi,$$
$$x_2 = \cosh t \sin \chi \sin \vartheta \sin \varphi, \qquad x_3 = \cosh t \sin \chi \cos \vartheta,$$
$$x_4 = \cosh t \cos \chi,$$

then the metric takes on the form

$$g = -dt^2 + \cosh^2 t\{d\chi^2 + \sin^2 \chi(d\vartheta^2 + \sin^2 \vartheta \, d\varphi^2)\} \qquad (4.3.17)$$

(see Figure 56a). The sections where $t =$ constant are Riemannian spaces with constant positive curvature and radius $\cosh t$; the universe first contracts, and then expands again. The geodesic vector field $dt$, however, is not unique; if intersections are taken with surfaces at 45° to the $x_0$-axis, then half of the hyperboloid is covered by the coordinates

$$\hat{t} = \ln(x_0 + x_4), \qquad \hat{x}_j = \frac{x_j}{x_0 + x_4}, \qquad j = 1, 2, 3$$

in which the metric has the form

$$g = -d\hat{t}^2 + e^{2\hat{t}}(d\hat{x}_1^2 + d\hat{x}_2^2 + d\hat{x}_3^2) \qquad (4.3.18)$$

(see Figure 56b). The intersections where $\hat{t}$ = constant are expanding Euclidean spaces, in the sense that the geodesics $\hat{x}_j$ = constant, $j = 1, 2, 3$, grow steadily farther apart.† There are a great many facets of the de Sitter universe; we shall even be able to find coordinates in which $g_{ik}$ does not depend on time at all (4.4.42) will show that it is even a static space. In order to survey the causal relationships better, it is convenient to map the whole space into a compact set, in what is known as a **Penrose diagram**. To this end, we write the $t$ of (4.3.17) as

$$t' = 2 \arctan(\exp t) - \frac{\pi}{2},$$

which makes the metric

$$g = \cosh^2(t)\,(-dt'^2 + d\chi^2 + \sin^2\chi\, d\Omega^2),$$
$$d\Omega^2 = d\vartheta^2 + \sin^2\vartheta\, d\varphi^2, \qquad 0 < \chi < \pi, \qquad -\frac{\pi}{2} < t' < \frac{\pi}{2}. \tag{4.3.19}$$

The conformal equivalence to Minkowski space is again evident if the coordinates

$$t + r = \tan\frac{t' + \chi}{2}, \qquad t - r = \tan\frac{t' - \chi}{2},$$
$$0 < \chi < \pi, \qquad -\pi + \chi < t' < \pi - \chi,$$

are used to turn the Minkowski metric into

$$g = \left[2\cos\frac{t' + \chi}{2}\cos\frac{t' - \chi}{2}\right]^{-2}(-dt'^2 + d\chi^2 + \sin^2\chi\, d\Omega^2). \tag{4.3.20}$$

Both the de Sitter universe and Minkowski space are mapped into a relatively compact part of $\mathbb{R} \times S^3$. The difference between the causal structures of the two spaces comes about because they cover different parts of the $(t', \chi)$ plane (Figure 57). In de Sitter space, timelike geodesics begin on the lines $t' = -\pi/2$, $0 < \chi < \pi$, and end at $t' = \pi/2$, $0 < \chi < \pi$. There are some that do not intersect the past of a given point $p$, and so an observer at $p$ would be unaware of them—they are "beyond the particle horizon." Conversely, the union of the past light-cones of a particle's trajectory does not fill the whole space; there are points that an observer on a geodesic would never see, "beyond the event horizon." This does not occur in Minkowski space, where a timelike geodesic begins at a point $(t', \chi) = (-\pi, 0)$ and ends at $(\pi, 0)$. Only an accelerated particle could emerge from within the line $(t, r) = (-\infty, \infty)$ of Figure 57, and in that case it is possible for two accelerated observers never to see each other (recall (I: 6.4.10; 2)).

† This is observable, since, for example, the proton and electron in a hydrogen atom do not move geodesically, but are electrically bound together, which keeps the Bohr radius from expanding with the geodesics.

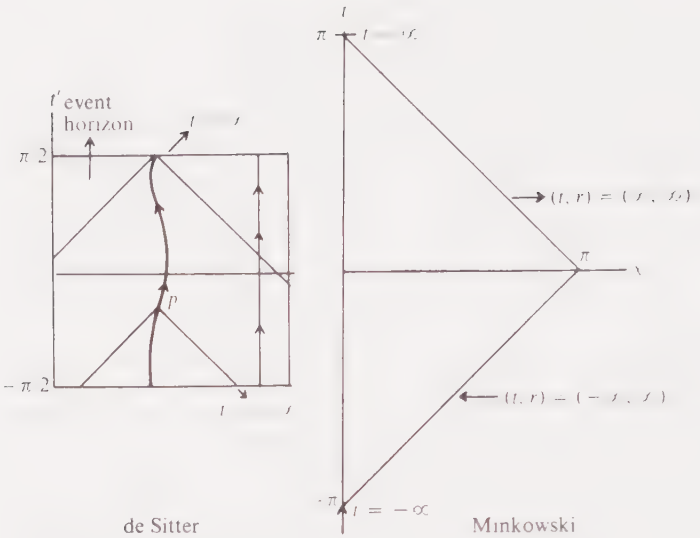



Figure 57 Penrose diagrams for flat and curved spaces

$K = -1$. **The Anti-de Sitter Universe** (4.3.21)

This can be represented as the hyperboloid

$$-x_0^2 + x_1^2 + x_2^2 + x_3^2 - x_4^2 = -1$$

in $\mathbb{R}^5$ with the metric

$$g = -dx_0^2 + dx_1^2 + dx_2^2 + dx_3^2 - dx_4^2$$

(Figure 58). We now observe that the intersection with $x_1 = x_2 = x_3 = 0$ is a closed, timelike geodesic. If the causal structure (cf. (I: 6.4.7)) is to be saved, it is necessary to pass to a covering surface, which can be mapped

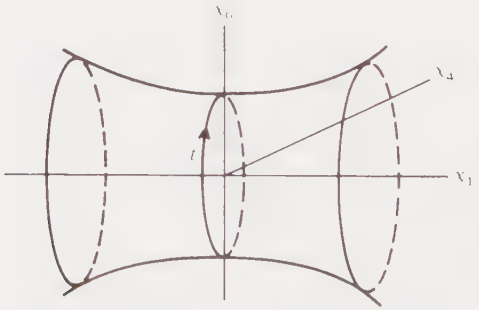


Figure 58 The anti-de Sitter universe

onto the region $\mathbb{R} \times \mathbb{R}^+ \times S^2$ for the variables $(t', \eta, \vartheta, \varphi)$ by setting

$$r = \sqrt{x_1^2 + x_2^2 + x_3^2} = \sinh \eta, \qquad x_0 = \cosh \eta \cos t',$$
$$x_4 = \cosh \eta \sin t'.$$

The metric is turned into

$$g = -\cosh^2 \eta \, dt'^2 + d\eta^2 + \sinh^2 \eta \, d\Omega^2. \tag{4.3.22}$$

Now the intersection where $t' = 0 = x_4$ is a Riemannian space of constant negative curvature. A sphere of radius $R$ in this space has the surface area $4\pi \sinh^2 R$, and proper time $ds = dt' \cosh R$ elapses much faster at large distances. To be able to compare this physical system with the earlier ones, we change to radial variables $\chi = 2 \arctan(\exp \eta) - \pi/2$. This makes the metric

$$g = \cosh^2 \eta(-dt'^2 + d\chi^2 + \sin^2 \chi \, d\Omega^2),$$
$$0 < \chi < \frac{\pi}{2}, \qquad -\infty < t' < \infty, \tag{4.3.23}$$

which is again of the form (4.3.19), except that the overall factor depends on $\chi$ rather than $t'$. The Penrose diagram is an infinite strip, which shows that in

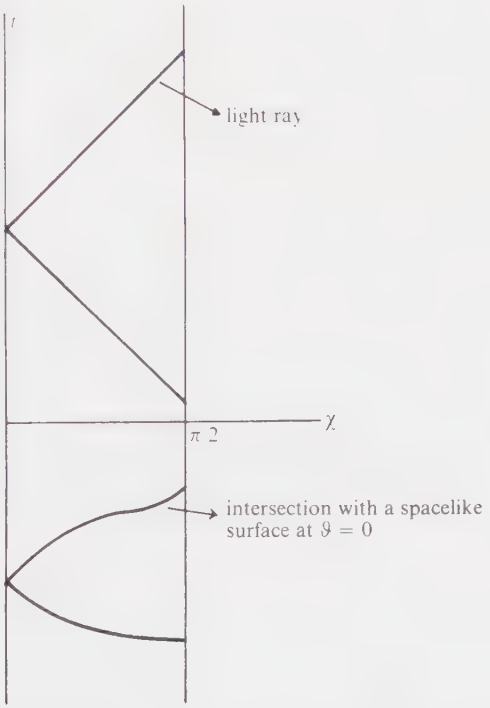



Figure 59 Penrose diagram for the covering surface of the anti-de Sitter universe

anti-de Sitter space there are no Cauchy surfaces at all. If we take the intersection with a spacelike surface at $\vartheta = 0$, then it must lie at an angle of less than 45° in the Penrose diagram. It is always possible to find a light ray along $\vartheta = 0$ which will never intersect this surface (Figure 59). Thus, in the covering surface, which has infinitely many sheets, time has the unusual property that for any infinite spacelike surface it is possible to find an event at a much later time having no causal connection with it.

**Einstein's Equations for Isotropic Spaces** (4.3.24)

If $R_{\alpha\beta} = Ke_{\alpha\beta}$, then $R_\beta = i_\alpha R^\alpha{}_\beta = 3Ke_\beta$, $R = i_\alpha R^\alpha = 12K$, and $R_\alpha - e_\alpha R/2 = -3Ke_\alpha$, so the energy-momentum tensor is $T_{\alpha\beta} = -\eta_{\alpha\beta} 3K/8\pi\kappa$. In the phenomenological description (3.1.25; 3), this corresponds to a fluid at rest with energy density = −pressure = $3K/8\pi\kappa$. A negative pressure is necessary to maintain the de Sitter universe ($K > 0$), and a negative energy is needed for the anti-de Sitter universe ($K < 0$), and of course the distributions must be homogeneous. Such unphysical $T_{\alpha\beta}$'s could be caused by

(a) a contribution $\Lambda$*1 in $\mathscr{L}$ (cf. (4.2.14; 2));
(b) the vacuum expectation value of the energy-momentum tensor of fields, which can be $\sim \simeq \eta_{\alpha\beta}$ for reasons of invariance [40];
(c) extra terms in Einstein's equations [31].

There is no very persuasive reason to believe in any of these suggestions, so, despite their esthetic appeal, maximally symmetric spaces are not the front-running candidates for models of the universe; in fact, the empirical evidence goes somewhat against them.

**Remarks** (4.3.25)

1. Einstein's equations control only the local behavior of space, and not its global structure. They do not say whether it is necessary to enclose the anti-de Sitter universe in a covering surface to save the causal structure, or whether it is permissible to destroy the causal structure of the de Sitter universe by identifying antipodal points.
2. Since we have discovered that quite simple spaces require drastic revisions of our familiar causality relationships, we have to be prepared for the worst when we couple fields to the energy-momentum tensor. With gravitation, these phenomena occur only on the cosmic scale, but similar, though strong, interactions for hadrons could wreak havoc on the small scale.

**Problems** (4.3.26)

1. What is the condition on the $g_{lm}$ of the natural basis so that $v$ is a Killing vector field?
2. Show that the Poisson bracket of $p_\alpha v^\alpha(x)$ with the Hamiltonian $g^{\alpha\beta}(x)p_\alpha p_\beta$ vanishes iff $v$ is a Killing vector field (cf. (I: 5.1.10; 2)).

3. Suppose there exists a Killing vector field v. Find a coordinate system in which the $g_{lm}$ of the natural basis are independent of one of the coordinates.

4. It is always true in two dimensions that $R_{ik} = K(x)e_{ik}(x)$ (cf. (4.1.43)). Why is $K$ not necessarily constant?

5. Let $e^0 = dt$ in an orthogonal basis. Show that $e^0$ is a geodesic vector field (see 4.1.41.3).

6. Calculate $d{}^*F^\alpha$ in the basis (4.3.9). What is the sign of the gravitational energy?

7. Calculate the integral $\int {}^*\mathbf{1}$ over a Riemannian space of constant positive curvature ($K = 1$).

## Solutions (4.3.27)

1. It is that $L_v\, dx^m = d(i_v\, dx^m) = dv^m = v^m{}_{,i}\, dx^i$ (cf. (I: 2.5.12; 5)); thus

$$0 = L_v g_{lm}\, dx^l\, dx^m = v^i g_{lm,i}\, dx^l\, dx^m + 2g_{lm} v^l{}_{,i}\, dx^i\, dx^m$$
$$\Rightarrow v^i g_{lm,i} + g_{im} v^i{}_{,l} + g_{il} v^i{}_{,m} = 0.$$

2. $\{p_\alpha v^\alpha(x),\ g^{\beta\gamma}(x) p_\beta p_\gamma\} = -2p_\alpha v^\alpha{}_{,\gamma} g^{\gamma\beta} p_\beta + v^\alpha g^{\beta\gamma}{}_{,\alpha} p_\beta p_\gamma = 0$ for Killing vector fields: $g_{\alpha\beta} g^{\beta\gamma} = \delta_\alpha{}^\gamma \Rightarrow g^{\beta\gamma}{}_{,\alpha} = -g^{\beta\sigma} g_{\sigma\rho,\alpha} g^{\rho\gamma}$, which brings us to the condition of Problem 1.

3. By Theorem (I: 2.3.12), it is always possible to find local coordinates for which the 1-component of v (in the natural basis) equals 1, and the others vanish. Then by Problem 1, $g_{lm,1} = 0$. If $v$ is timelike, this coordinate can be treated as time, making the metric constant and $v = \partial_t$.

4. There are no 3-forms in two dimensions, and hence there is no Bianchi identity.

5. $0 = de^0 = \omega^0{}_k \wedge e^k$ implies that $\omega^0{}_k \sim e^k$, for $k = 1, 2, 3$. Hence (cf. (4.1.11) and (4.1.17)),

$$D_{e_0} e^0 = -e^k(\omega^0{}_k | e) = 0.$$

6. The basis is orthogonal, hence $\varepsilon_{\alpha\beta\gamma\delta} = -\varepsilon^{\alpha\beta\gamma\delta}$, and

$$-\tfrac{1}{2}\varepsilon^{\alpha\beta\gamma\delta}\omega_{\beta\gamma} \wedge e_\delta = -\frac{K}{2}\varepsilon^{\alpha\beta\gamma\delta} x_\beta e_{\gamma\delta};$$

$$-\tfrac{1}{2}\varepsilon^{\alpha\beta\gamma\delta}\, d(\omega_{\beta\gamma} \wedge e_\delta) = -\frac{K}{2}\varepsilon^{\alpha\beta\gamma\delta}\left[e_{\beta\gamma,\delta}\left(1 + \frac{Kr^2}{4}\right) - Kx_\beta x^\sigma e_{\sigma\gamma\delta}\right]$$
$$= -\frac{K}{2}\left[3!{}^*e^\alpha\left(1 + \frac{Kr^2}{4}\right) - Kx_\beta x^\sigma 2! e_\sigma \wedge {}^*e^{\alpha\beta}\right]$$
$$= -3K{}^*e^\alpha + \frac{K^2 r^2}{4}{}^*e^\alpha - K^2 x^\alpha x_\beta {}^*e^\beta = 8\pi\kappa({}^*t^\alpha + {}^*\mathcal{T}^\alpha).$$

The term linear in $K$ is the same as the right side of Einstein's version, and the parts $\sim K^2$ are the gravitational contribution $\sim\omega \wedge \omega$, which is always negative for the energy density $T_{00}$: $-K^2((|\mathbf{x}|^2 + 3t^2)/4$. It is necessarily negative if $K > 0$, since the integral of the energy density over a compact space must be zero (1.3.35; 2), and the part $\sim K$ is positive.

7. $\int *\mathbf{1} = \int d^m x/(1 + r^2/4)^m = S_m \int_0^\infty dr\, r^{m-1}/(1 + r^2/4)^m$, where $S_m = 2\pi^{m/2}/\Gamma(m/2)$ is the surface area of the $m$-sphere. If $\beta = (1 + r^2/4)^{-1}$, then, recalling that $\Gamma(m)/\Gamma(m/2) = (2^{m-1}/\sqrt{\pi})\Gamma(\frac{1}{2}(m + 1))$,

$$\int *\mathbf{1} = S_m 2^{m-1} \int_0^\infty d\beta\, \beta^{m-2} \left(\frac{1}{\beta} - 1\right)^{(m-2)/2} = S_m 2^{m-1} \frac{\Gamma(m^2/2)}{\Gamma(m)}$$

$$= \frac{2\pi^{(m+1)/2}}{\Gamma\left(\dfrac{m+1}{2}\right)} = S_{m+1}.$$

Since the space is isometric to the $m + 1$-sphere, the calculation of the volume is correct.

## 4.4 Spaces with Maximally Symmetric Submanifolds

*The nonlinearity complicates Einstein's equations so much that the general solution lies beyond human capabilities. Explicit solutions can be written down only if the space is of sufficiently high symmetry.*

If the symmetry of a maximally symmetric space is reduced, the variety of possible curvature forms becomes great enough to conceivably correspond to physically acceptable energy-momentum currents. It would, however, lead too far afield if we tried to classify all the possibilities exhaustively, so instead our plan will be to investigate the physically relevant metrics that come up when the symmetry is reduced in successive stages.

### Spaces with Six Killing Vector Fields (4.4.1)

The interesting case is the **Friedmann universe**, with six spacelike Killing vector fields, generating a group isomorphic to O(4). The trajectories of a point under the action of the group form a spacelike submanifold with six Killing vector fields, i.e., a three-dimensional Riemannian space of constant curvature. It is convenient to choose co-moving coordinates, for which the geodesic vector field furnishes the coordinate lines $x$ = constant perpendicular to this space, and the proper time on these geodesic lines is the time-coordinate $t$ (cf. (4.3.26; 5)). Writing $r^2 = |\mathbf{x}|^2$, the metric $g$ is of the type of a

### Robertson–Walker Metric (4.4.2)

$$g = -dt^2 + R(t)^2 \frac{|d\mathbf{x}|^2}{(1 + Kr^2/4)^2}.$$

**Remarks** (4.4.3)

1. $R(t)$ is an as yet unspecified function of time, something like $\cosh(t)$ in de Sitter space (see (4.3.17), and note that $|d\mathbf{x}|^2(1 + Kr^2/4)^{-1}$ is $d\chi^2 + \sin^2\chi\, d\Omega^2$ in the coordinates used there), or like $\exp(t)$ for $K = 0$ (see (4.3.18)). If $K > 0$, then the submanifold $t =$ constant has the finite volume $2\pi^2 R^3(t)/K^{3/2}$ (see (4.3.26; 7)).
2. If a new time-variable $t'$ such that $dt'/dt = 1/R(t)$ is introduced as in (4.3.17), then $g$ becomes conformally equivalent to that of de Sitter space, and consequently of Minkowski space,

$$g = R(t)^2\left(-dt'^2 + \frac{|d\mathbf{x}|^2}{(1 + Kr^2/4)^2}\right)$$

(cf. (4.3.19)). It frequently happens that $t'$ takes values only in a finite interval $t_0 < t < t_1$, as in de Sitter space. In that case the causality relationships turn out to be similar to those discussed in (4.3.16), and in particular there are particle and event horizons.

**The Curvature Forms of the Friedmann Universe** (4.4.4)

Let Greek indices run from 0 to 3 and Roman ones from 1 to 3. If we write the orthogonal basis as

$$e^\alpha = \left(dt, \frac{R(t)\,dx^a}{1 + Kr^2/4}\right),$$

then

$$de^\alpha = \left(0, \frac{\dot R}{R}e^{0a} - \frac{K}{2R}x_b e^{ba}\right),$$

and therefore

$$\omega^a{}_0 = \omega^0{}_a = \frac{\dot R}{R}e^a, \qquad \omega_{ab} = \frac{K}{2R}(x_a e_b - x_b e_a). \tag{4.4.5}$$

Consequently,

$$d\omega^{0a} = \frac{\ddot R}{R}e^{0a} - \frac{K\dot R}{2R^2}x_b e^{ba},$$

$$d\omega^{ab} = \frac{K}{R^2}e^{ab}\left(1 + \frac{Kr^2}{4}\right) + \frac{K^2}{4R^2}(x^a x_c e^{bc} - x^b x_c e^{ac})$$

and a similar calculation to that of (4.3.8) leads to

$$R^{0a} = \frac{\ddot R}{R}e^{0a}, \qquad R^{ab} = \frac{K + \dot R^2}{R^2}e^{ab}. \tag{4.4.6}$$

The contracted quantities become

$$R^0 = 3\frac{\ddot{R}}{R}e^0, \qquad R^a = \left(\frac{\ddot{R}}{R} + 2\frac{K + \dot{R}^2}{R^2}\right)e^a,$$
$$R = 6\left(\frac{K + \dot{R}^2(t)}{R^2(t)} + \frac{\ddot{R}(t)}{R(t)}\right). \tag{4.4.7}$$

**Remarks** (4.4.8)

1. If $R(t) =$ constant, then the curvature is constant only in spatial directions,

$$R^{ab} = \frac{K}{R^2}e^{ab}, \qquad R^{0a} = 0.$$

   The time-independence of $R$ gives rise to an $R^{0a}$ and contributes to $R^{ab}$.
2. A comparison of (4.4.7) and (4.1.44; 3) reveals that the Weyl forms are now zero, as required by the conformal equivalence to Minkowski space.

**Einstein's Equations in the Classical Form** (4.4.9)

In order to satisfy (4.2.20(a)), the energy-momentum forms of matter must be $e^\alpha \times$ (some function of $t$), because of (4.4.7). Therefore the energy-momentum tensor of matter is necessarily diagonal, and in the spirit of the phenomenological description (3.1.25; 3) we set $T_{00} = \rho =$ energy density and $T_{jj} = p =$ pressure. Einstein's equations then imply that

$$3\frac{\dot{R}^2 + K}{R^2} = 8\pi\kappa\rho$$
$$-\frac{2\ddot{R}}{R} - \frac{\dot{R}^2 + K}{R^2} = 8\pi\kappa p. \tag{4.4.10}$$

**Remarks** (4.4.11)

1. The Bianchi identity (4.2.32) relates $\rho$ and $p$ to $R$, and of course the same relationship follows from (4.4.10). It implies that

$$d^*\mathscr{T}_0 = d(\rho^* e^0) = d\rho \wedge {}^*e^0 + \rho\, d^*e^0$$
$$= -\omega^0{}_j \wedge {}^*\mathscr{T}^j = -p\omega^0{}_j \wedge {}^*e^j,$$

   and therefore

$$d\rho \wedge {}^*e^0 = (\rho - p)\omega^0{}_j \wedge {}^*e^j, \text{ i.e., } \dot{\rho} = 3\frac{\dot{R}}{R}(\rho - p).$$

   In the form

$$p = -\frac{\frac{d}{dt}(\rho R^3)}{\frac{d}{dt}R^3}$$

this has the interpretation that pressure = −(rate of change of energy)/(rate of change of volume). It is noteworthy that gravity does not appear in the total energy in co-moving coordinates.

2. The static situation $\dot{R} = 0$ requires a negative pressure if $K > 0$, as in de Sitter space, and a negative energy if $K < 0$. This originally induced Einstein to include the cosmological term $\Lambda^*1$ in his action principle. Friedmann later discovered the solution bearing his name, and the modern tendency is to accept the dynamical equations as is. In order to illustrate their significance better, let us examine

**Einstein's Equations in Landau and Lifshitz's Form** (4.4.12)

In (4.2.20(a)) and (4.2.21(b)) the energy-momentum forms for matter and gravitation were represented as exterior differentials of 2-forms. Since it is not yet apparent what the gravitational energy is, let us track it down. To this end we calculate the restriction of the exterior differential of the 2-form

$$-\tfrac{1}{2}\varepsilon^{0bcd}\omega_{bc} \wedge e_d = -\frac{K}{2R}\varepsilon^{0bcd}x_b e_{cd},$$

$$-\tfrac{1}{2}\varepsilon^{0bcd}\, d(\omega_{bc} \wedge e_d)\Big|_{t=\text{const}} = -\frac{K}{2R^2}\varepsilon^{0bcd}\left[e_{bcd}\left(1+\frac{Kr^2}{4}\right) - Kx_b x_m e_{mcd}\right]\Bigg|_{t=\text{const}}$$

$$= \frac{3K - K^2r^2/4}{R^2} *e^0\Big|_{t=\text{const}} = 8\pi\kappa(*\mathscr{T}^0 + *t^0) \tag{4.4.13}$$

to $t = \text{constant}$.

**Remarks** (4.4.14)

1. Equation (4.4.13) states that

$$8\pi\kappa \cdot \text{the total energy density} = \frac{3K - K^2r^2/4}{R^2}$$

so $8\pi\kappa \cdot$ the gravitational energy density equals $8\pi\kappa \cdot$(total energy density $- \rho$) $= -(K^2r^2 + 12\dot{R}^2)/4R^2$. In the static situation ($\dot{R} = 0$) this is the energy density of a homogeneous mass distribution $\rho$ according to Newton's theory, constructed just as in electrodynamics, but with the other sign and $4\pi\kappa \to e^2$. If we chose units where $R = 1$, then

$$\Delta V = e\rho = \frac{3K}{2e}, \qquad V = \frac{Kr^2}{4e},$$

$$\nabla V = \frac{\mathbf{x}K}{2e}, \qquad -\tfrac{1}{2}|\nabla V|^2 = -\frac{1}{2e^2}\frac{K^2r^2}{4},$$

and thus, as promised in §1.3, $-K^2r^2/4R^2$ corresponds exactly to $8\pi\kappa \cdot$ the Newtonian gravitational energy in the most nearly Cartesian coordinates. In the dynamical case, there is also a contribution $\dot{R}^2$.

2. The gravitational energy is exactly large enough so that if $K > 0$, the integral of the total energy over the whole space is zero:

$$3K \int_0^\infty \frac{r^2\, dr}{(1 + Kr^2/4)^3} = \frac{K^2}{4} \int_0^\infty \frac{r^4\, dr}{(1 + Kr^2/4)^3}.$$

3. If we write the first of Equations (4.4.10) as

$$\frac{\dot{R}^2}{2} - 4\pi\kappa\rho \frac{R^2}{3} = -\frac{K}{2} = \text{constant},$$

then it has the form of the conservation equation for the energy of a (non-relativistic) particle with coordinate $R$ and speed $\dot{R}$. The kinetic energy plus the potential energy is constant, where the latter is taken as the potential energy on the surface of a ball of radius $R$ and homogeneous density $\rho$:

$$-V(r) = \frac{4\pi R^3 \rho}{3r}\,\Theta(r - R) + \frac{4\pi R^2}{2}\left(-1 + \frac{r^2}{3R^2}\right)\Theta(R - r).$$

Note that $V(0) - V(R)$ is smaller than $V(R) - V(\infty)$ by a factor of 2.

In the following section, we shall return to (4.4.10) when we study the collapse of stars, and we shall solve them for selected pressure-density ratios. A detailed discussion of their importance for cosmology may be found in [8], [9], [10].

**Spaces with Five Killing Vector Fields** (4.4.15)

In (I: §5.8) we learned that there are in general 5 constants of the motion in the field of a gravitational wave, which are linear in the momentum components. This motion accordingly allows 5 Killing vector fields, and the invariance group they generate is isomorphic to the invariance group of a gravitational wave, and consequently (I: 5.8.4: 1) to that of an electromagnetic wave (I: 5.5.3). The metric given in (I: 5.8.1) is a special case of what we examined in (4.2.13), where it depended on only one coordinate $u$. The interesting situation is where discontinuities are allowed, in which $du = n_\alpha e^\alpha$, $0 = n_\alpha n^\alpha \equiv n^2$. If the coordinates are chosen so that

$$n = (-1, 1, 0, 0), \qquad u = x - t, \tag{4.4.16}$$

then the metric (I: 5.8.1), with the required invariance structure, is of the form

$$g = -dt^2 + dx^2 + p(u)^2\, dy^2 + q(u)^2\, dz^2. \tag{4.4.17}$$

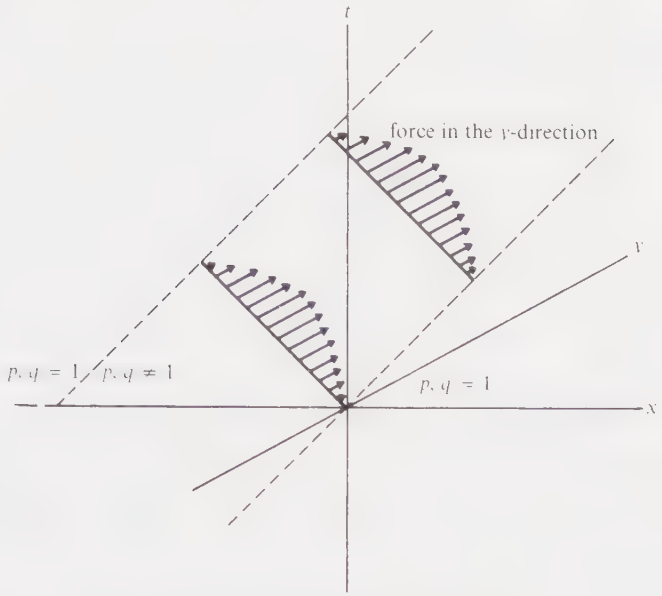



Figure 60 Schematic drawing of a gravitational pulse

**Remarks** (4.4.18)

1. The form of $g$ is that of a metric in a co-moving coordinate system (cf. (4.3.26; 5)). Therefore the coordinate lines $\mathbf{x} =$ constant are possible particle trajectories.
2. The gravitational field described by $g$ is a kind of transverse wave, which alters the distance perpendicular to the direction of propagation between particle trajectories. If $p - 1$ and $q - 1$ have compact support, then the pulse looks schematically as shown in Figure 60. For example, in the solution given in (I: 5.8.7) $p \leq 1$ and $q \leq 1$, so the trajectories draw coser together in the $y$-direction and spread apart in the $z$-direction, as with a quadrupole field. This effect would not be measured by measuring rods, as they would be stretched in the same way. However, the deformation would be observable by measuring the time-delay of a reflected light signal.

**The Curvature Forms** (4.4.19)

If the orthogonal basis

$$e^\alpha = (dt, dx, p\,dy, q\,dz),$$
$$de^\alpha = (0, 0, p'\,du \wedge dy, q'\,du \wedge dz)$$

is used, then the affine connections become

$$\omega_{\alpha\beta} = S_\alpha n_\beta - S_\beta n_\alpha, \qquad S_\alpha = (0, 0, p'\,dy, q'\,dz),$$

as in (4.2.13). Since $n_\alpha S^\alpha = 0 = n_\alpha n^\alpha$, we find that $\omega_{\alpha\beta} \wedge \omega^\beta{}_\gamma = 0$ for all $\alpha$ and $\beta$, and the curvature forms are

$$\begin{aligned} R_{\alpha\beta} &= d\omega_{\alpha\beta} = du \wedge (S'_\alpha n_\beta - S'_\beta n_\alpha), \\ S'_\alpha &= (0, 0, p'' dy, q'' dz), \end{aligned} \tag{4.4.20}$$

with the contractions

$$\begin{aligned} R_\beta &= i_\alpha R^\alpha{}_\beta = -n_\beta du\, i_\alpha S'^\alpha = -n_\beta du\left(\frac{p''}{p} + \frac{q''}{q}\right), \\ R &= 0. \end{aligned} \tag{4.4.21}$$

**Remarks** (4.4.22)

1. In two dimensions, $g = -dt^2 + p^2(t)dy$, the curvature is $p'' dt \wedge dy$, and thus $R_{\alpha\beta}$ are precisely the curvatures of the corresponding two-dimensional surfaces.
2. The Weyl forms are not of necessity zero, and the space need not be conformally flat. However, if $p'' = q'' = 0$, then it is always flat (cf. Remark (I: 5.8.8; 2)).

**Einstein's Equations in the Classical Form** (4.4.23)

By Equations (4.4.21),

$$R_\alpha - \tfrac{1}{2} e_\alpha R = -n_\alpha du\left(\frac{p''}{p} + \frac{q''}{q}\right),$$

and thus an energy current $\sim du$ could well be a source of the gravitational pulse, as long as it is accompanied by an equally large current of the 1-component of the momentum while the other components vanish. Such an energy-momentum current could be produced by an electromagnetic wave. If

$$F = f(u) \wedge du \quad \text{with} \quad \langle f \mid du\rangle = \langle du \mid du\rangle = 0,$$

then $F \wedge {}^*F = 0$, and the energy-momentum forms (1.3.22) are

$$\begin{aligned} \mathcal{T}_\alpha &= {}^*((i_\alpha F) \wedge {}^*F) = -i_{i_\alpha F} F = -(f_\alpha i_{du} - n_\alpha i_f) f \wedge du \\ &= n_\alpha du \langle f \mid f\rangle. \end{aligned}$$

Therefore Einstein's equations imply that

$$\frac{p''}{p} + \frac{q''}{q} + 8\pi\kappa \langle f \mid f\rangle = 0.$$

To solve this equation, set $p = L\exp(\beta)$ and $q = L\exp(-\beta)$. Then

$$\frac{L''}{L} + \beta'^2 + 4\pi\kappa\langle f \mid f\rangle = 0, \tag{4.4.24}$$

$$\beta(u) = \int_0^u du' \left(-\frac{L''(u')}{L(u')} - 4\pi\kappa\langle f(u') \mid f(u')\rangle\right)^{1/2} \tag{4.4.25}$$

**Remarks** (4.4.26)

1. In the approximation linear in $\kappa$,

$$L(u) = 1 - 4\pi\kappa \int_0^u du' \int_0^{u'} du'' \langle f(u'') | f(u'') \rangle,$$

while $\beta$ remains arbitrary to first order. If the equations are homogeneous ($f = 0$), this provides us with a solution that could be used for $\varphi^{\text{in}}$ in (4.2.18), because to first order, $\beta = \varphi_{22} = -\varphi_{33}$, and the other $\varphi$'s are zero, so that $0 = \varphi_\alpha{}^\alpha = \varphi_{\beta\alpha}{}^{,\alpha} = \varphi_{\alpha\alpha,\rho}{}^{,\rho}$.
2. If $L > 0$, then it is a concave function, because $\langle f | f \rangle > 0$, and $f$ and $\beta'$ contribute in similar ways to the curvature $L''$ of the function $L(u)$. In this situation, the trajectories of particles are focused in the $y - z$-plane. This is an effect of the gravitational field produced by the electromagnetic or gravitational wave.
3. If $L'$ is ever negative, then $L$ must sooner or later have a zero. This singularity in the metric might not be a genuine one, but may only indicate that the gravitational wave has disrupted the coordinate system. The space might appear as Minkowski space in some other chart (cf. (I: 5.8.8; 2)), as soon as the wave has passed.

**Einstein's Equations in Landau and Lifshitz's Form** (4.4.27)

It remains to find out how well the interpretation (4.4.26; 2) of $\beta'^2$ as an energy density accords with the formulations (4.2.21(b)). In the latter formulation gravity contributes

$$\ell_\alpha = \frac{-1}{8\pi\kappa}\frac{1}{2}\omega_{\beta\gamma} \wedge \omega_{\alpha\rho} \wedge {}^*e^{\beta\gamma\rho}$$

to the energy, and the other summand is zero. Substitution from (4.4.19) yields

$$\begin{aligned} 8\pi\kappa\ell_\alpha &= {}^*(S_\beta n_\gamma \wedge S_\rho n_\alpha \wedge {}^*e^{\rho\beta\gamma}) \\ &= -n_\alpha i_{S_\beta} i_{S_\rho}(e^{\rho\beta} \wedge du) = -n_\alpha\, du\, 2\frac{p'q'}{pq} \qquad (4.4.28) \\ &= 2n_\alpha\, du\left(\beta'^2 - \left(\frac{L'}{L}\right)^2\right), \end{aligned}$$

so that an additional negative term $-(L'/L)^2$ occurs along with $\beta'^2$.

**Remarks** (4.4.29)

1. In the linear approximation (4.4.26; 1), $L' = 0$, $2\dot{\beta} = \dot{g}_{22} = -\dot{g}_{33}$, and the result is that

$$\text{energy density} = \text{momentum density} = \frac{(\dot{g}_{22} - \dot{g}_{33})^2}{64\pi\kappa}.$$

2. Provided that $L$ has a slowly varying amplitude in comparison with $\beta$, (4.2.21(b)) states that $\dot{\beta}^2$ creates gravity like any other kind of energy, and that the energy of a gravitational wave is positive.
3. Fictitious energies associated with fictitious forces also appear in (4.4.27); their origin is that the $\ell_\alpha$ do not vanish even in flat space ($p'' = q'' = 0$).
4. The speeding-up of a double star with a short period seems to be consistent with the energy loss due to gravitational radiation as calculated with this formula.

**Spaces with 4 Killing Vector Fields** (4.4.30)

We shall consider the spaces that are the counterpart to the problem of a central force in mechanics. The energy and angular momentum will correspond to the operations of time-displacement and rotations that leave $g$ invariant. In the polar coordinates for $M = \mathbb{R} \times \mathbb{R}^+ \times S^2$, the $g_{\alpha\beta}$ depend only on $r = |\mathbf{x}|$, and the metric can be written as

$$g = -dt^2 \exp(2a(r)) + dr^2 \exp(2b(r)) + r^2(d\vartheta^2 + \sin^2 \vartheta \, d\varphi^2).$$

**Remarks** (4.4.31)

1. If Einstein's equations in vacuo hold, then it can be shown that the existence of the timelike Killing vector field follows simply from the spherical symmetry (Birkhoff's theorem, Problem 5).
2. In co-moving coordinates, the $g_{\alpha\beta}$ are in general time-dependent.

**The Curvature Forms** (4.4.32)

In the orthogonal basis

$$e^\alpha = (e^a \, dt, e^b \, dr, r \, d\vartheta, r \sin \vartheta \, d\varphi),$$

$$de^\alpha = (a' e^a \, dr \wedge dt, 0, dr \wedge d\vartheta, \sin \vartheta \, dr \wedge d\varphi + r \cos \vartheta \, d\vartheta \wedge d\varphi)$$

the affine connections turn out to be

$$\omega^{\alpha}{}_{\beta} = \begin{array}{|c|c|c|} \hline e^{a-b}a'\,dt & 0 & 0 \\ \hline & -e^{-b}\,d\vartheta & -e^{-b}\sin\vartheta\,d\varphi \\ \hline & & -\cos\vartheta\,d\varphi \\ \hline \end{array} \tag{4.4.33}$$

(since $\omega_{\alpha\beta} = -\omega_{\beta\alpha}$, we write them only for $\alpha < \beta$). They make

$d\omega^{\alpha}{}_{\beta} =$

| $e^{a-b}(a'' + a'(a' - b'))dr \wedge dt$ | 0 | 0 |
|---|---|---|
| | $e^{-b}b'\,dr \wedge d\vartheta$ | $e^{-b}(-\cos\vartheta\,d\vartheta + \sin\vartheta b'\,dr) \wedge d\varphi$ |
| | | $\sin\vartheta\,d\vartheta \wedge d\varphi$ |

$$\omega^{\alpha}{}_{\gamma} \wedge \omega^{\gamma}{}_{\beta} = \begin{array}{|c|c|c|} \hline 0 & -e^{a-2b}a'\,dt \wedge d\vartheta & -e^{a-2b}a'\sin\vartheta\,dt \wedge d\vartheta \\ \hline & 0 & e^{-b}\cos\vartheta\,d\vartheta \wedge d\varphi \\ \hline & & e^{-2b}\sin\vartheta\,d\vartheta \wedge d\varphi \\ \hline \end{array}.$$

The term $\sim d\vartheta \wedge d\varphi$ cancels out of $d\omega^{1}{}_{3}$, and the $R^{\alpha}{}_{\beta}$ become proportional to $e^{\alpha}{}_{\beta}$:

$$R^{\alpha}{}_{\beta} = \begin{array}{|c|c|c|} \hline (a'b' - a'' - a'^{2})e^{-2b}e^{0}{}_{1} & -\dfrac{a'e^{2b}}{r}e^{0}{}_{2} & -\dfrac{a'e^{2b}}{r}e^{0}{}_{3} \\ \hline & \dfrac{b'e^{2b}}{r}e^{1}{}_{2} & \dfrac{b'e^{2b}}{r}e^{1}{}_{3} \\ \hline & & \dfrac{(1-e^{2b})}{r^{2}}e^{2}{}_{3} \\ \hline \end{array}. \tag{4.4.34}$$

**Einstein's Equations** (4.4.35)

Since the remaining symmetry still suffices to make $R^{\alpha}{}_{\beta}$ of the form $R^{\alpha}{}_{\beta} = K^{\alpha}{}_{\beta} e^{\alpha}{}_{\beta}$ (no sum), the energy-momentum forms $\mathscr{T}^{\alpha}$ (letting this embrace everything coupled to gravity) must likewise be $\sim e^{\alpha}$. The coefficients are $-\sum_{\substack{\beta<\gamma \\ \beta\neq\alpha,\,\gamma\neq\alpha}} K_{\beta\gamma}$, and depend only on $r$:

$$
\begin{aligned}
\left(-\frac{1-e^{-2b}}{r^2} - 2b'\frac{e^{-2b}}{r}\right)e^0 &= 8\pi\kappa\mathscr{T}^0, \\
\left(-\frac{1-e^{-2b}}{r^2} + 2a'\frac{e^{-2b}}{r}\right)e^1 &= 8\pi\kappa\mathscr{T}^1, \\
\left(a'' + a'^2 - a'b' + \frac{a'-b'}{r}\right)e^{-2b}e^2 &= 8\pi\kappa\mathscr{T}^2, \\
\left(a'' + a'^2 - a'b' + \frac{a'-b'}{r}\right)e^{-2b}e^3 &= 8\pi\kappa\mathscr{T}^3.
\end{aligned}
\tag{4.4.36}
$$

**Remarks** (4.4.37)

1. Because of the spherical symmetry, there is an invariance under $2 \leftrightarrow 3$, and more specifically $\mathscr{T}^2$ and $\mathscr{T}^3$ have the same factor in front of $e^2$ and respectively $e^3$.
2. $\mathscr{T}^{\alpha}$ is of the form $c^{\alpha}e^{\alpha}$ (no sum), and the contracted Bianchi identity (4.2.32) subjects the coefficients $c^{\alpha}$ to the equation

$$dc^{\alpha} \wedge {}^*e^{\alpha} = \sum_{\beta} \omega^{\alpha}{}_{\beta} \wedge {}^*e^{\beta}(c^{\alpha} - c^{\beta}).$$

3. The $\mathscr{T}^{\alpha}$ are written with the basis of (4.4.32), and thus the $\mathscr{T}^j$, $j = 1, 2, 3$, are obtained from Cartesian energy-momentum forms by local rotations (cf. (1.3.26)). If, for instance, $\mathscr{T}^j = pe^j$ in the Cartesian basis, then it is also true in this basis.

**Special Cases** (4.4.38)

1. If we make the phenomenological assumption that

$$\mathscr{T}_{\alpha} = (\rho e^0, pe^j)$$

(cf. (3.1.25; 3)), where $\rho$ and $p$ are not too singular and decrease sufficiently fast as $r \to \infty$, then the first of Equations (4.4.36) is solved by

$$e^{-2b} = 1 - \frac{8\pi\kappa}{r}\int_0^r dr'\, r'^2\rho(r') \equiv 1 - \frac{2\kappa M(r)}{r}, \tag{4.4.39}$$

and the second one determines $a$ once $b$ is known:

$$a = -b - 4\pi\kappa \int_r^\infty dr'\, r' e^{2b(r')}(\rho(r') + p(r')). \tag{4.4.40}$$

The last two identical equations relate $\rho$ to $p$, and are equivalent to requiring the contracted Bianchi identity. This subject will be pursued in the following section, for the $\mathcal{T}^\alpha$ treated here it is satisfied.

If it happens that $\rho(r) = p(r) = 0$ for all $r > r_1$, then the metric in the region where $r > r_1$ is the

**Schwarzschild Metric** (4.4.41)

$$g = -\left(1 - \frac{r_0}{r}\right)dt^2 + \frac{dr^2}{1 - \frac{r_0}{r}} + r^2\, d\Omega^2, \qquad r_0 = 8\pi\kappa \int_0^\infty r^2\, dr\, \rho(r).$$

On the other hand, if $\rho = \text{constant} = -p$, then we return to the situation of (4.3.23) and obtain (cf. Problem 4) a

**Static Form of the de Sitter Metric** (4.4.42)

$$g = -(1 - Kr^2)dt^2 + \frac{dr^2}{1 - Kr^2} + r^2\, d\Omega^2, \qquad K = \frac{8\pi\kappa}{3}\rho.$$

2. Equations (4.4.36) allow the pressure in the radial direction to differ from the pressure in the $\vartheta$ and $\varphi$ directions. This could happen for the Coulomb field of a point particle, for which the energy-momentum forms can be calculated as

$$\mathcal{T}^\alpha = \frac{e^2}{2r^4}[e^0, -e^1, e^2, e^3]. \tag{4.4.43}$$

If we set $\exp(2a) = \exp(-2b) = \psi(r)/r$ in (4.4.36), then Einstein's equations read:

$$\left(\frac{1 - \psi'}{r^2}e^0, \frac{-1 + \psi'}{r^2}e^1, \frac{\psi''}{2r}e^2, \frac{\psi''}{2r}e^3\right) = 8\pi\kappa\mathcal{T}_\alpha.$$

If $\psi = r - r_0 + 4\pi\kappa e^2/r$, then this simply reproduces (4.4.43), and the resulting metric is called the

**Reissner–Nordstrøm Metric** (4.4.44)

$$g = -\left(1 - \frac{r_0}{r} + \frac{4\pi\kappa e^2}{r^2}\right)dt^2 + \frac{dr^2}{1 - \frac{r_0}{r} + \frac{4\pi\kappa e^2}{r^2}} + r^2\, d\Omega^2.$$

**Remarks** (4.4.45)

1. In the linear approximation, and with $|p| \ll \rho$, Equation (4.4.40) becomes

$$a = -\frac{g_{00} - 1}{2} = -4\pi\kappa\left[\frac{1}{r}\int_0^r dr'\,\rho(r')r'^2 + \int_r^\infty dr'\,\rho(r')r'\right],$$

which is the Newtonian potential of the spherically symmetric energy density $\rho$, as it must be on account of (4.2.30; 2).

2. The pressure contributes to (4.4.40) as the density contributes to $g_{00}$. Hence the negative pressure $p = -\rho$ of the de Sitter universe in fact makes $a =$ —the Newtonian potential, because for constant densities,

$$\frac{1}{r}\int_0^r r'^2\,dr'\,\rho = -\frac{1}{2}\int_r^R r'\,dr'\,\rho + \text{constant}.$$

3. If $\rho$ is more singular than $r^{-3}$ at $r = 0$, then one can write $\int_r^\infty$ instead of $-\int_0^r$ in (4.4.39). This is the case with the Reissner solution (4.4.44), so the positive field energy contributes with a reversed sign to the gravitational potential. The interpretation is that $M = M(\infty)$ represents the total energy, and the potential

$$-\frac{M\kappa}{r} + \frac{4\pi e^2\kappa}{2r^2}$$

shows that as one approaches the origin, part of the energy density is left behind, and the potential is effectively decreased compared with its asymptotic value, $-M\kappa/r$. If $M < \infty$, then it follows that the "naked mass" at the origin must be $-\infty$, since the electromagnetic mass

$$\frac{4\pi}{2}\int_0^\infty \frac{dr}{r^4}$$

is divergent. This has the paradoxical consequence that gravity is repulsive at short distances. Once again, the infinite electrostatic self-energy of a point charge is causing trouble.

4. The basis of (4.4.32) is less suitable for a discussion of gravitational energy using the version (4.2.21(b)) of Einstein's equations, since it is possible to simulate a gravitational $\ell_x$ even in flat space, with polar coordinates. In Problem 6 the gravitational energy is discussed in the maximally Cartesian coordinates (see (I: 5.7.17; 4)). It turns out that as long as there is asymptotically a Schwarzschild metric, the total energy including the gravitational energy is $M$. Note that

$$\int_{t=0} {}^*\mathscr{T}^0 = 4\pi\int_0^\infty dr\,r^2 e^{b(r)}\rho(r) \neq M = 4\pi\int_0^\infty dr\,r^2\rho(r).$$

5. In Problem 7 the energy density of gravitation is calculated as $-\kappa M^2\,8\pi r^4$ in these coordinates. This is equal to the negative of the energy density

$(e\ 4\pi r^2)^2\ 2$ of a Coulomb field, where $e^2$ is replaced with $4\pi\kappa M^2$, analogously to Remark (4.4.14; 1). Since $M$ is the integral of the total energy density, there can be an everywhere regular solution with $M > 0$ only if $\rho > 0$ counterbalances the negative gravitational energy. The increase of the Schwarzschild potential compared with the Newtonian potential at small $r$ can be interpreted as the field produced by the negative gravitational energy.

**Properties of Spherically Symmetric Fields** (4.4.46)

1. The Geometric Interpretation of the Spatial Metric. The restriction of the metric to a plane passing through the origin,

$$g_{\substack{t=\text{const}\\ \varphi=\text{const}}} = e^{2b}\,dr^2 + r^2\,d\vartheta^2,$$

is the metric on a surface of rotation in $\mathbb{R}^3$. If it is written in cylindrical coordinates as $z(r)$, then

$$dz^2 + dr^2 + r^2\,d\vartheta^2 = dr^2(1 + z'^2) + r^2\,d\vartheta^2 = e^{2b}\,dr^2 + r^2\,d\vartheta^2,$$

or, using (4.4.39),

$$z' = \sqrt{1 - e^{2b}} = \sqrt{\frac{2\kappa M(r)}{r - 2\kappa M(r)}},$$

$$z(r) = \int_0^r dr' \sqrt{\frac{2\kappa M(r')}{r' - 2\kappa M(r')}}.$$

As a consequence, the Schwarzschild metric $M(r) = \text{constant}$ gives the intersection $\varphi = \text{constant}$ the geometry of a paraboloid of revolution,

$$z(r) = \sqrt{4r_0}\sqrt{r - r_0}$$

(Figure 61). The metric is singular at $r = 2\kappa M(r)$. The paraboloid of the Schwarzschild metric can be extended beyond that point, but, if so, $r$ is no longer a monotonic function of $z$.

2. The Causal Structure. The Schwarzschild metric was extended beyond $r = r_0$ in (I: 5.7.2; 5), with the aid of the coordinates

$$\begin{aligned} u &= \sqrt{\frac{r}{r_0} - 1}\exp\left(\frac{r}{2r_0}\right)\cosh\left(\frac{t}{2r_0}\right) \\ v &= \sqrt{\frac{r}{r_0} - 1}\exp\left(\frac{r}{2r_0}\right)\sinh\left(\frac{t}{2r_0}\right), \qquad u^2 - v^2 > -1, \end{aligned} \tag{4.4.47}$$

in which

$$g = 4r_0^3\exp\left(-\frac{r}{r_0}\right)\frac{du^2 - dv^2}{r} + r^2\,d\Omega^2. \tag{4.4.48}$$

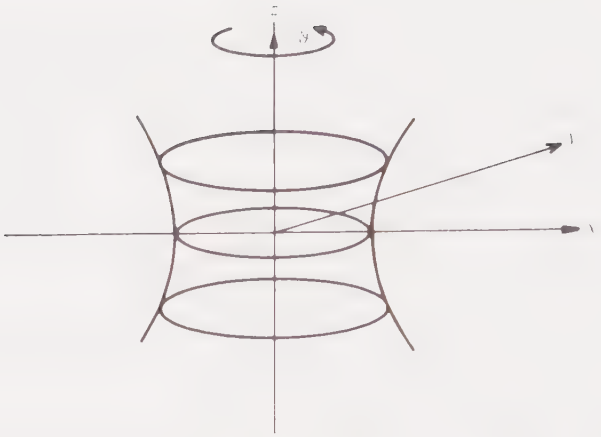


Figure 61 A choice of the metric on the surface $t$ = constant, $\varphi$ = constant

There remains a singularity at $r = 0$, which is now a spacelike hypersurface $u^2 - v^2 = -1$. The quickest way to understand the causal relationships is to draw the Penrose diagram that results from using the coordinates

$$v + u = \tan\left(\frac{\psi + \xi}{2}\right), \qquad v - u = \tan\left(\frac{\psi - \xi}{2}\right),$$
$$-\pi < \psi \pm \xi < \pi, \qquad -\frac{\pi}{2} < \psi < \frac{\pi}{2}. \tag{4.4.49}$$

The metric

$$g = A^2(-d\psi^2 + d\xi^2 + R^2\, d\Omega^2),$$
$$A = \frac{r_0}{\sqrt{r}}\, e^{-r/2r_0} \cos^{-1}\tfrac{1}{2}(\psi + \xi)\cos^{-1}\tfrac{1}{2}(\psi + \xi),$$
$$R = \frac{r}{A},$$

reveals that timelike lines run in the $\psi$-$\xi$-plane at angles of at least 45°, and radial light rays run at 45°. The region covered by the new time and radial coordinates looks as shown below. Since the boundary contains the spacelike piece where $r = 0$, there is a horizon. Although $r = r_0$ is not a singularity, it is the event horizon for all trajectories that remain in Region I, where $r > r_0$. Regions II and III are invisible from Region I, which is itself invisible from Regions III and IV. Although nothing exceptional happens locally at $r = r_0$, the surface $r = r_0$ has a global significance.

The Reissner metric (4.4.44) becomes for some $r \in (0, \infty)$ singular if $4\pi\kappa e^2 < M^2\kappa^2$. For hadrons this inequality is far from being satisfied since in natural units $e^2 = \frac{1}{137} \geq (\kappa 4\pi)M^2 \equiv$ (Planck length/Compton wave length)$^2 = (10^{-33}\text{cm}\ 10^{-14}\text{ cm})^2$. Thus no horizon prevents people from

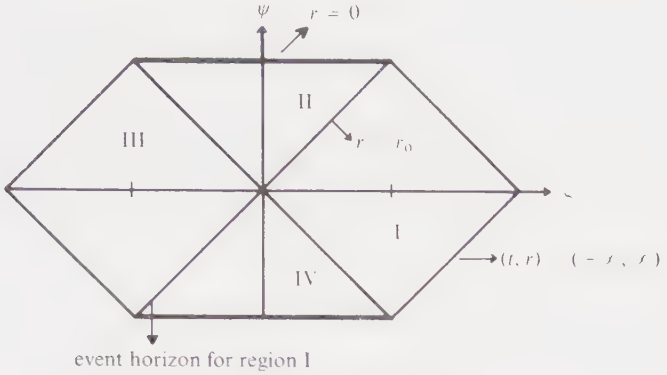



Figure 62 Penrose diagram for the Schwarzschild solution

starting at the "naked singularity" at $r = 0$. The "cosmic censorship hypothesis" conjectures that naked singularities do not develop in reality. Indeed $4\pi e^2 \geq M^2\kappa$ means that the Coulomb repulsion of charged matter would be stronger than the gravitational attraction thus preventing the collapse to a singularity. If $M^2\kappa^2 > 4\pi\kappa e^2$, then the singularity of (4.4.44) at small $r$ lies only in the choice of coordinates; with other coordinates it would be possible to continue to $r = 0$. In that case, the repulsive nature of gravity makes $r = 0$ a time-like line. The appropriate Penrose diagram, Figure 63, thus extends in the timelike direction to infinity, as with the anti-de Sitter universe ([9], p. 921). Hence there are again no global Cauchy surfaces, but instead there is a bizarre possibility that one might crawl through the wormhole bounded by $r = 0$ into another universe just like ours (I′ in Figure 63).

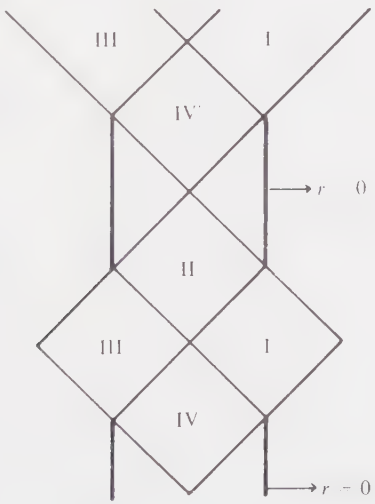



Figure 63 Penrose diagram for the Reissner solution

3. Singularities. Now that the singularity at $r = r_0$ has successfully been removed from the Schwarzschild metric, the question arises of whether the singularity at $r = 0$ is genuine. It is not necessarily significant that the $g_{ik}$ are infinite there, since these quantities depend on the coordinates. However, the invariant $*(R_{\alpha\beta} \wedge *R^{\alpha\beta})$ goes as $r^{-6}$, and grows without bound as $r \to 0$ (Problem 8). This would not be possible at a regular singular point, so we conclude that the Schwarzschild metric can not be extended regularly across the region $u^2 - v^2 > -1$.

**Problems** (4.4.50)

1. Construct the five Killing vector fields of the metric (4.4.17) (cf. (I: 5.8.3)).

2. In the linear approximation, the metric at large distances

$$g_{\alpha\beta} = \eta_{\alpha\beta} + \frac{4\kappa}{r} \int d^3x' (T_{\alpha\beta} - \tfrac{1}{2} T \eta_{\alpha\beta})_{t-|x-x'|}$$

looks like the field of a plane wave. Using (4.4.29; 1), calculate the energy radiated in the 1-direction, and express it in terms of the quadrupole tensor

$$D_{ab} = \int d^3x \; T_{00}(3x_a x_b - \delta_{ab}|\mathbf{x}|^2).$$

3. Calculate the $\mathscr{T}_\alpha$ of (4.4.43) for $F = (e/r^2)e^{01}$. Verify that $i_{e^\alpha} \mathscr{T}_\alpha = 0$.

4. Reexpress (4.4.42) in the form (4.3.9).

5. Prove Birkhoff's theorem: If the $a$ and $b$ of (4.4.30) also depend on time, and Einstein's equations hold with $\mathscr{T}_\alpha = 0$, then there exists a time-coordinate $t'$ such that $g$ is of the form (4.4.30) with $a$ and $b$ independent of time. The metric $g$ must then be of the form (4.4.41).

6. Use the Schwarzschild metric in the form

$$g = -h^2(r)dt^2 + f^2(r)|d\mathbf{x}|^2, \qquad f = \left(1 + \frac{\kappa M}{2r}\right)^2,$$

$$h = \left(1 - \frac{\kappa M}{2r}\right)\left(1 + \frac{\kappa M}{2r}\right)^{-1},$$

(cf. (I: 5.7.17; 4)) and the version (4.2.21(b)) to identify the total energy

$$\int_{N_3} *(\mathscr{T}^0 + \mathscr{U}^0)$$

with $M$. Use a ball of radius $R \to \infty$ for $N_3$.

7. Calculate the density of the total energy as in (4.4.45; 4) with the $\omega$'s of Problem 6.

8. Calculate $*(R_{\alpha\beta} \wedge *R^{\alpha\beta})$ for the Schwarzschild metric, and check that it is unbounded at $r = 0$.

**Solutions** (4.4.51)

1. The fields with components

$$v^i = (0, 0, 1, 0), (0, 0, 0, 1), (1, 1, 0, 0), \left(z, z, 0, \int_0^{t-x} \frac{du}{q^2(u)}\right), \left(y, y, \int_0^{t-x} \frac{du}{p^2(u)}, 0\right)$$

satisfy

$$v^i g_{im,i} + g_{im} v^i{}_{,l} + g_{il} v^i{}_{,m} = 0,$$

and are therefore Killing vector fields according to (4.3.27; 1).

2. From the continuity equation of the linear approximation, $T^{\alpha\beta}{}_{,\beta} = 0$, it follows that

$$\frac{\partial^2}{\partial t^2} T_{00} = \nabla_a \nabla_b T_{ab}, \quad \text{i.e.,} \int d^3x\, T_{ab} = \frac{1}{2} \frac{\partial^2}{\partial t^2} \int d^3x\, x_a x_b T_{00}.$$

Therefore

$$\dot{g}_{ab} = \frac{2\kappa}{r} \frac{\partial^3}{\partial t^3} \int d^3x\, x_a x_b T_{00} + \eta_{ab} c.$$

The contribution $\sim \eta_{ab}$ is irrelevant, because we require only the square of the difference of the eigenvalues of the 2-3 subspace $(\dot{g}_{33} - \dot{g}_{22})^2 + 4\dot{g}_{32}^2$. Substitution of $D_{ab}$ yields

$$\mathscr{T}_{10} = \left(\frac{\kappa}{144\pi r^2}\right)[(\ddot{D}_{33} - \ddot{D}_{22})^2 + 4\ddot{D}_{23}^2]$$

(cf. (1.1.13) and [32, §104]).

3.
$$*F = -\frac{e}{r^2} e^{23}, \qquad \mathscr{T}_\alpha = \tfrac{1}{2}[-i_{i_\alpha *F} *F - i_{i_\alpha F} F],$$

$$\mathscr{T}_0 = -\tfrac{1}{2} \frac{e}{r^2} i_{e_1} \frac{e}{r^2} e^{01} = \frac{e^2}{r^4} e^0,$$

$$\mathscr{T}_1 = \tfrac{1}{2} \frac{e}{r^2} i_{e_0} \frac{e}{r^2} e^{01} = -\frac{e^2}{r^4} e^1,$$

$$\mathscr{T}_2 = -\tfrac{1}{2} \frac{e}{r^2} i_{e_3} \frac{e}{r^2} e^{23} = \frac{e^2}{r^4} e^2,$$

$$\mathscr{T}_3 = \tfrac{1}{2} \frac{e}{r^3} i_{e_2} \frac{e}{r^2} e^{23} = \frac{e^2}{r^4} e^3.$$

4. Suppose $K = 1$, and introduce the coordinates

$$x_4 = (1 - r^2)^{1/2} \cosh t \quad \text{and} \quad x_5 = (1 - r^2)^{1/2} \sinh t$$

on the surface where $x_1^2 + x_2^2 + x_3^2 + x_4^2 - x_5^2 = 1$. Then

$$g = |d\mathbf{x}|^2 + dx_4^2 - dx_5^2 = |d\mathbf{x}|^2 - (1 - r^2)dt^2 + \frac{r^2\, dr^2}{1 - r^2}.$$

If $K = -1$, then take the coordinates

$$x_4 = (1 + r^2)\cos t \quad \text{and} \quad x_5 = (1 + r^2)\sin t$$

on the surface where $-|\mathbf{x}|^2 + x_4^2 + x_5^2 = 1$.

5. If $a$ and $b$ depend on time, then the only immediate change is $\omega^0{}_1$, by $\exp(b - a)\dot{b}\, dr$. This produces the following extra terms in the $8\pi\kappa\mathscr{T}^\alpha$ of (4.4.35):

$$\alpha = 0\colon 2\dot{b}\frac{e^{-a-b}}{r}e^1$$

$$\alpha = 1\colon -2\dot{b}\frac{e^{-a-b}}{r}e^0$$

$$\alpha = 2, 3\colon -e^{-2a}(\ddot{b} + \dot{b}^2 - \dot{a}\dot{b})e^\alpha.$$

If $\alpha = 0$, then we also find that $\dot{b} = 0$ and hence $\exp(-2b) = 1 - r_0/r$. And if $\alpha = 1$, then $a' = -b'$, and thus $\exp(2a) = (1 - r_0/r)f^2(t)$. With the variables $dt' = f(t)dt$, we have the Schwarzschild metric. No new conditions result if $\alpha = 2$ or 3.

6. $$e^\alpha = (h\, dt,\ f\, d\mathbf{x}), \qquad \omega^{0j} = \frac{h'}{f}\frac{x^j}{r}dt, \qquad \omega^{jk} = \frac{f'}{fr}(x^k\, dx^j - x^j\, dx^k),$$

$$\begin{aligned} 8\pi\kappa\int_{N_3}(*\mathscr{T}^0 + *\mathscr{t}^0) &= -\frac{1}{2}\varepsilon^{0bcd}\int_{\partial N_3}\omega_{bc}\wedge e_d \\ &= -\varepsilon^{0bcd}\int_{\partial N_3}\frac{f'}{r}x^c\, dx^b\wedge dx^d \\ &= -\lim_{R\to\infty}\int_{r=R}d\Omega\, 2f'r^2 = 8\pi\kappa M. \end{aligned}$$

In the above equations, $d\Omega$ is the element of solid angle, and we have recalled that $-\varepsilon^{0123} = \varepsilon_{0123} = 1$.

7. $8\pi\kappa(*\mathscr{T}^0 + *\mathscr{t}^0) = -\frac{1}{2}\varepsilon^{0bcd}\, d(\omega_{bc}\wedge e_d) = -\varepsilon^{0bcd}$.

$$\begin{aligned} d\left(\frac{f'}{r}x^c\right)dx^b\wedge dx^d &= -dx^1\wedge dx^2\wedge dx^3\left[6\frac{f'}{r} + 2r\left(\frac{f'}{r}\right)'\right] \\ &= -\frac{(\kappa M)^2}{r^4}dx^1\wedge dx^2\wedge dx^3. \end{aligned}$$

8. $R^{\alpha\beta} = c^{\alpha\beta}e^{\alpha\beta}$ (no sum);

$$c^{\alpha\beta} = \frac{\kappa M}{r^3}\begin{array}{|c|c|c|}\hline 2 & -1 & -1 \\ \hline & -1 & -1 \\ \hline & & 2 \\ \hline\end{array},$$

$$*(R_{\alpha\beta}\wedge *R^{\alpha\beta}) = \sum_{\alpha,\beta}c_{\alpha\beta}^2 = \frac{24(\kappa M)^2}{r^6}.$$

## 4.5 The Life and Death of Stars

*Gravity differs from other interactions by having a very small coupling constant, and by being universal. For cosmic bodies, the latter property makes the action of gravity sum constructively to such an extent that it dominates all other interactions.*

**The Orders of Magnitude** (4.5.1)

The gravitational energy of $N$ protons (mass $m$) in a volume $V$ is on the order of

$$E_G \sim -\frac{\kappa(Nm)^2}{V^{1/3}} = -\kappa m^2 N^{2/3} N\rho^{1/3}, \qquad \rho = \frac{N}{V}.$$

Although the Coulomb interaction is unimaginably stronger than this, $e^2 \sim 10^{36}\,\kappa m^2$, it is neutralized in normal matter, so that the electrical energy per particle is $\sim -e^2$ (the distance between nearest neighbors). This distance is $\sim \rho^{-1/3}$, so the total electrostatic energy is

$$E_e \sim -e^2 N\rho^{1/3} = \frac{e^2}{\kappa m^2 N^{2/3}} E_G. \tag{4.5.2}$$

We see that if $N \sim (e^2/\kappa m^2)^{3/2} \sim 10^{54}$, then gravity starts to dominate the electrical forces. The mass of Jupiter is about that of $10^{54}$ protons, which is the point at which the Newtonian potential supplants the Coulomb potential as the determiner of the structure. In a larger body, gravity crushes the atoms together, and the matter turns into a highly compressed plasma.

The Fermi energy, which is the origin of the solidity of matter, is (the number of electrons) $\times$ (the nearest-neighbor distance)$^{-2}$ $\times$ (the electron mass)$^{-1}$, in natural units ($\hbar = c = 1$):

$$E_F \sim \frac{N\rho^{2/3}}{m_e}. \tag{4.5.3}$$

The density $\rho$ of an object adjusts so as to minimize the total energy. Whereas for the Coulombic energy (4.5.2) this makes the density independent of $N$,

$$\rho \sim (e^2 m_e)^3 = (\text{Bohr radius})^{-3}, \tag{4.5.4}$$

in the case of gravitation objects containing more particles are smaller:

$$\rho^{1/3} \sim m^2 \kappa N^{2/3} m_e, \qquad V \sim N^{-1} m_e^{-3} (\kappa m^2)^{-3}. \tag{4.5.5}$$

However, as soon as the separation between nearest neighbors is on the order of magnitude of the Compton wavelength, $\rho^{1/3} \sim m_e$, the relativistic energy $|p|$ is to be used in calculating $E_F$ instead of $|p|^2/2m_e$, and (4.5.3) is replaced by

$$E_F \sim N\rho^{1/3}. \tag{4.5.6}$$

The gravitational energy consequently dominates the Fermi energy when $\kappa m^2 N^{2/3} > 1 \Rightarrow N > (\kappa m^2)^{-3/2} \sim 10^{57}$, i.e., when the mass is somewhat greater than the mass of the sun; and the minimum of the total energy is attained when $\rho = \infty$ and $v = 0$. After that point, there is a process in nature that dramatically controls what happens. The rate of energy loss from stars is normally rather slow – one photon takes several million years to escape from the interior of the sun – but sufficiently energetic electrons can create neutrinos by inverse beta decay $e^- + p \to \nu + n$, which, as they feel no strong interaction, leave the star immediately. This makes the transition to states of lower energy proceed at a much higher rate, and in a matter of seconds the star collapses to a neutron star, of nuclear density. Hence the energy released is on the order of the kinetic energy of neutrons at this density, about 10 MeV per particle, and thus as much energy is emitted as in the normal thermonuclear reactions, but much more rapidly. That is why it is assumed that the catastrophe just described is what takes place in a supernova, for which a single star may radiate with the brilliance of a whole galaxy for a week. The energy released would be the same, because a galaxy has typically $10^{10}$ stars, and normally a star takes $10^9$ years $\sim 10^{10}$ weeks to burn up all its nuclear fuel.

This line of reasoning makes use of a naive, Newtonian picture of gravity, and it is interesting to see how it changes in Einstein's theory, with the help of the material developed in the preceding section. It might be hoped that a sufficiently great pressure could counteract the gravitational attraction and render the stars stable. This is not necessarily the case, however, because in the relativistic theory pressure can also produce gravity, which can aggravate the situation.

Recall, in the spirit of the phenomenological description of the energy and momentum of matter,

$$\mathcal{T}_0 = -\rho e_0, \qquad \mathcal{T}_j = p e_j, \qquad j = 1, 2, 3, \tag{4.5.7}$$

that the energy density $\rho$ and pressure $p$ can not be chosen completely arbitrarily, because of the contracted Bianchi identity (4.2.32) connecting them; for the special form (4.5.7) it requires that

$$dp \wedge {}^*e^1 = \omega^1{}_0 \wedge {}^*e^0(p + \rho) \tag{4.5.8}$$

(see (4.4.37; 2) with $\alpha = 1$ and $\omega^1{}_\beta$ only nonzero if $\beta = 0$).

**Remarks** (4.5.9)

1. We shall later be primarily interested in the static, spherically symmetric case. Then both sides of the equation in (4.4.37; 2) vanish for $\alpha \neq 1$, with the $e$'s and $\omega$'s of (4.4.30); hence (4.5.8) contains all the information of (4.2.32).
2. Since we have earlier expressed the metric in terms of $\rho$ and $p$, Equation (4.5.8) creates a relationship between $\rho$ and $p$, which must be satisfied in

order to have static equilibrium. If an equation of state is known for $\rho$ and $p$, then there can be a static state only at the density distribution for which (4.5.8) agrees with the equation of state.

Taking the $\omega$ of (4.4.33),

$$dp \wedge {}^*e^1 = \exp(-b)\left(\frac{\partial p}{\partial r}\right){}^*\mathbf{1} = -\exp(-b)a'(\rho + p){}^*\mathbf{1}, \qquad (4.5.10)$$

and according to (4.4.39) and (4.4.40) (with $\kappa M(r) \to M(r)$),

$$\begin{aligned} a' &= -b' + 4\pi\kappa r(\rho + p)e^{2b} \\ &= \left(1 - \frac{2M(r)}{r}\right)^{-1} \cdot \left[-4\pi\kappa r\rho + \frac{M(r)}{r^2} + 4\pi\kappa r(\rho + p)\right] \end{aligned}$$

When this is substituted into (4.5.10), there results the

**Tolman–Oppenheimer–Volkoff Equation** (4.5.11)

$$-\frac{dp}{dr} = \frac{(\rho + p)[M(r) + 4\pi\kappa p r^3]}{r(r - 2M(r))}.$$

**Remarks** (4.5.12)

1. Of course, this also follows from (4.4.36), but the Bianchi identity does the trick without the extraneous information of (4.4.36).
2. Equation (4.5.11) generalizes the nonrelativistic fact that

$$-\frac{\partial p}{\partial r} = \frac{\rho M(r)}{r^2}.$$

The increase of the pressure for decreasing $r$ is intensified by the following relativistic effects:

(a) There is an additional term $\sim p$ in $M(r)$, since pressure also produces gravity;
(b) It is necessary to add $p$ to $\rho$, since the gravitational force also acts on $p$;
(c) Gravity increases faster than $\sim 1/r^2$ as $r \to 0$.

We saw at the outset that large, gravitating masses lose their stability in the special theory of relativity, because a relativistic electron gas is not as stiff as a nonrelativistic one, and does not stand firm against gravity. The general relativistic situation is even more precarious, because the solidity of matter also fails to help. In order to see this, we integrate (4.5.11) for the most extreme equation of state, viz., that of incompressible matter, which can not be squashed to arbitrarily high density. If $\rho$ = constant and we require the boundary condition $p(R) = 0$, where $R$ is the radius of the star, then in

dimensionless variables we find

$$x = r\sqrt{\frac{8\pi\kappa\rho}{3}}, \qquad x_0 = R\sqrt{\frac{8\pi\kappa\rho}{3}} = \sqrt{\frac{r_0}{R}}, \tag{4.5.13}$$

$$p(x) = \rho\frac{\sqrt{1-x^2}-\sqrt{1-x_0^2}}{3\sqrt{1-x_0^2}-\sqrt{1-x^2}}$$

(Problem 1). As a consequence, we can read off the

**Maximal Pressure in Homogeneous Stars** (4.5.14)

$$p(0) = \rho\frac{1-\sqrt{1-r_0/R}}{3\sqrt{1-r_0/R}-1}.$$

**Consequences** (4.5.15)

1. Whereas $p(0)$ goes as $\rho r_0\, 4R$ for stars of homogeneous densities whose radii are much larger than the Schwarzschild radius, and thus $p(0)$ is normally much less than $\rho$, if $R \to r_0$ it increases rapidly and becomes infinite at $R = 9r_0/8$.
2. The pressure in matter comes from the electrons, while the protons give rise to the energy density. The relative orders of magnitude are that $p/\rho \sim$ (electron speed $v)^2 \times$ (electron mass)/(proton mass) $\sim v^2 \cdot 10^{-3}$, so that the pressure in the center of a star like the sun, with $R \sim 10^5 r_0$, requires electrons to be moving at $\sim\frac{1}{10}$ the speed of light. The electrons must be relativistic in stars of the same mass but hundreds of times smaller (white and black dwarfs), and the situation becomes critical.

The next question to answer is how Einstein's theory affects the naive expression (4.5.1) for the gravitational energy. In §4.4 we saw that

$$\frac{r_0}{2\kappa} = 4\pi\int_0^\infty r^2\,dr\,\rho(r) \equiv M$$

is the total energy of the system, while

$$\int \rho^* e^0 = 4\pi\int_0^\infty r^2\,dr\,\rho(r)\left(1-\frac{2M(r)}{r}\right)^{-1/2} \tag{4.5.16}$$

equals the total energy of the matter alone. If $\rho$ = constant, then Equation

(4.5.16) can be evaluated easily, and there results the

**Gravitational Energy of a Homogeneous Star** (4.5.17)

$$E_G = M - \int \rho^* e^0 = M\left(1 - \frac{3}{2}\left[\arcsin\sqrt{\frac{r_0}{R}} - \sqrt{\frac{r_0}{R}}\sqrt{1 - \frac{r_0}{R}}\right]\left(\frac{R}{r_0}\right)^{3/2}\right);$$

if

$$R \gg r_0 : E_G = -\frac{3}{5}\frac{\kappa M^2}{R}$$

$$R = r_0 : E_G = -M\left(\frac{3\pi}{4} - 1\right).$$

When the density is small, this reduces to the Newtonian self-energy of a ball of uniform density, and as $R \to r_0$ this formula is of the same order of magnitude, but its numerical factor is somewhat different. At the limit of stability $R = 9r_0/8$, it reads $E_G = -0.37M$.

If the pressure called for in (4.5.11) can not be provided, then it is not possible to have static equilibrium, and the star collapses. In order to pursue this drama analytically, let us consider only stars of uniform pressure and density distributions. The Friedmann solution (4.4.2) applies in the interior, while in the exterior the free Einstein equations hold, for which, according to Birkhoff's theorem (4.4.50; 5) the only available solution with spherical symmetry is the Schwarzschild solution. The problem of matching the solutions will be discussed later, after we study the dynamics in the interior.

The point of deparature is Equations (4.4.10), which will be used in the form

$$p = -\frac{\frac{d}{dt}(\rho R^3)}{\frac{d}{dt}R^3} \tag{4.5.18}$$

$$\frac{\dot{R}^2}{2} - \kappa\frac{4\pi}{3}\frac{\rho R^3}{R} = -\frac{K}{2}. \tag{4.5.19}$$

We solve them first for the extremal equation of state, $p = 0$. In normal matter, $p$ is always much less than $\rho$: the greatest pressure is that of massless particles, $p = \rho/3$. When either $p = \rho/3$ or $p = \rho$, analytic solutions can be written down (Problem 3), and we shall later figure out the qualitative behavior for all $p > 0$, which is generally similar.

**Solutions with** $p = 0$ (4.5.20)

Equation (4.5.18) implies that $M = 4\pi\rho R^3/3 =$ constant, making (4.5.19) of the form of the energy of the radial Kepler motion, with no angular momentum. This equation was integrated in (I: §4.2), and the solution is most conveniently written in the form of Kepler's equation (I: 4.2.24; 7).

We can identify the variables used there as

$$m = 1, \qquad E = -\frac{K}{2}, \qquad \alpha = -\kappa M \Rightarrow a = \frac{\kappa M}{K}, \qquad \varepsilon = 1,$$

and must distinguish three cases:

(a) $K > 0$

$$R = \frac{\kappa M}{K}(1 - \cos u), \qquad t - t_0 = \frac{\kappa M}{K^{3/2}}(u - \sin u) \tag{4.5.21}$$

(b) $K = 0$

$$R = (t - t_0)^{2/3}\left(\frac{9\kappa M}{2}\right)^{1/3} \tag{4.5.22}$$

(c) $K < 0$

$$R = \frac{\kappa M}{|K|}(\cosh u - 1), \qquad t - t_0 = \frac{\kappa M}{|K|^{3/2}}(\sinh u - u). \tag{4.5.23}$$

**Remarks** (4.5.24)

1. The case $K > 0$ in (4.5.19) corresponds to a negative energy in the Kepler problem. Accordingly, $R$ equals zero when $u = 0$ ($t = t_0$) and when $u = 2\pi$ ($t = t_0 + 2\pi\kappa M/K^{3/2}$). In this case Kepler's equations (4.5.21) are the parametric representation of a cycloid, specifying how time elapses during a free fall into the center:

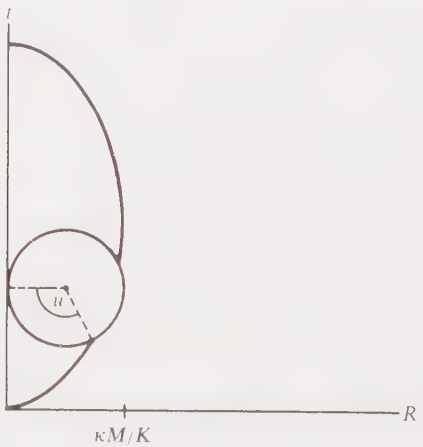



Figure 64 Cycloid for $R(t)$

2. Case (a) describes a big bang at $t = t_0$, which was so weak that the particles flying along the geodesics $x = \text{constant}$ are eventually stopped by gravity, turn around, and all will eventually crash together again. In case (b), on the other hand, the initial velocity is high enough to send the particles to infinity. The space $t = \text{constant}$ is not compact, but in fact simply $\mathbb{R}^3$.
3. In case (c) the particles retain some kinetic energy when they reach infinity, and the space $t = \text{constant}$ has negative curvature.
4. If $t = t_0$, then $R = 0$, and we learn from (4.4.7) that the metric has a genuine singularity at that point, because the curvature scalar approaches infinity.

These results are probably not too surprising, since matter without pressure or angular momentum would be expected to fall into the center unless it has a large enough initial outward radial velocity. A positive pressure changes nothing, because the extra gravity it produces actually favors the collapse. The reason can be seen formally in (4.4.10), by which a positive $p$ contributes negatively to $\ddot{R}$, thus increasing the concavity of the function $t \to R(t)$. In that case, $R(t)$ must approach zero if the initial slope is too small. To be mathematically precise, let us state these thoughts as a

**Criterion for Collapse of the Friedmann Universe** (4.5.25)

*Let* $\dot{R}(0)^2 < K > 0$ *and* $p \geq 0$. *Then* $R(t)$ *vanishes for some*

$$t < \frac{2R(0)}{\sqrt{K} - \dot{R}(0)},$$

*in which* $R(0)$ *is connected with* $\dot{R}(0)$ *by* $R(0)^2 = 3(K + \dot{R}(0))/8\pi\kappa\rho(0)$, *according to* (4.4.10).

**Proof**

Let us write the second equation of (4.4.10) as

$$\ddot{R}(t) = -\frac{\dot{R}(t)^2}{2R(t)} - \frac{K}{2R(t)} - \frac{R(t)8\pi\kappa p(t)}{2} \leq -\frac{K}{2R(t)}.$$

Then

$$R(t) = R(0) + t\dot{R}(0) + \int_0^t dt' \int_0^t dt'' \, \ddot{R}(t'')$$

$$\leq R(0) + t\dot{R}(0) - \frac{Kt^2}{4a} = R(0) + \frac{\dot{R}(0)^2 a}{K} - \frac{K}{4a}\left(t - \frac{2\dot{R}(0)a}{K}\right)^2,$$

$$\text{where } a \equiv \sup_t R(t).$$

If $\dot{R}(0)^2 < K$, then $a \leq R(0) + a\dot{R}(0)^2/K$ implies

$$a \leq R(0)\left(1 - \frac{\dot{R}(0)^2}{K}\right)^{-1},$$

and the zero for $R(t)$ happens before

$$t_0 = \frac{2\dot{R}(0)a}{K} + \sqrt{\frac{4aR(0)}{K} + \frac{4a^2\dot{R}(0)^2}{K^2}}, \quad \text{as} \quad R(0) + t_0\dot{R}(0) = \frac{Kt_0^2}{4a}.$$

The bound on $a$ implies (4.5.25) □

**Remarks** (4.5.26)

1. If $p = 0$, then the condition $\dot{R}(0)^2 < K$ corresponds to the statement for the equivalent Kepler problem that the kinetic energy is less than minus the total energy. This obvious criterion preventing escape is valid for all $p \geq 0$.
2. The time $t$ is that of a co-moving coordinate system, and thus the space collapses to a point within a finite proper time for freely falling observers.

Finally, we construct a solution of Einstein's equations, which describes gravitational collapse. The physical picture of what takes place is as follows: If, after having exhausted its nuclear fuel, a star has shrunk down so far that the Fermi energy of the electrons has risen above the threshold for inverse beta decay $e^- + p \to \nu + n$, then the greater part of the matter is turned into neutrons. Since the star is supported against collapse mainly by the Fermi pressure of the electrons, it suddenly gives way. Thus the model would be that of a star in static equilibrium, whose pressure at some time is suddenly reduced to zero. The solution of Einstein's equations before that time is as in (4.4.30). Afterwards, the solution in the interior is (4.4.2), and in the exterior it is the Schwarzschild metric. We now need to show that the solutions can be joined smoothly at the surface to satisfy Einstein's equations with $\rho$ = constant inside and 0 outside, and $p = 0$. Since the surface of the star falls freely, its radius in the co-falling coordinates (4.4.2) is $r = a$ = constant. For simplicity we use units in which $r_0 = 8\pi\kappa a^3\rho/3 = 1$ and consider the case $K = 0$. This makes the motion parabolic, with the surface of the star infinitely large at the beginning. Similarly, the solution with $K > 0$ is a Friedmann space in the interior, matched to a Schwarzschild metric. If $p > 0$, the calculation becomes much more complicated, because it can not be constant inside the star, as otherwise there would be an infinite pressure gradient at the surface. However, the essential features are not greatly altered if $p > 0$ [36].

In order to join (4.4.30) to (4.4.2), we have to express the two metrics in the same coordinates. For this reason, we write the

**Schwarzschild Metric in Co-Falling Coordinates** (4.5.27)

It is convenient to introduce the coordinates $(\tau, \tilde{r})$ in place of $(t, r)$, where $\tau$ is the proper time for radial parabolic motion, and $\tilde{r}$ is $r$ at the time $t = 0$. Since the speed approaches zero asymptotically,

$$p^0 = \left(\frac{dt}{d\tau}\right)\frac{r-1}{r} = 1,$$

and thus

$$-1 = \frac{r}{r-1}\left(-1 + \left(\frac{dr}{d\tau}\right)^2\right) \Rightarrow \frac{dr}{d\tau} = -\frac{1}{\sqrt{r}} \Rightarrow \tau = \frac{2}{3}(\tilde{r}^{3/2} - r^{3/2}).$$

Consequently,

$$\frac{dt}{dr} = \frac{d\tau}{dr}\left(1 + \frac{1}{r-1}\right) \Rightarrow t = \tau - 2\sqrt{r} + \ln\frac{\sqrt{r}+1}{\sqrt{r}-1}.$$

This puts the metric in the normal form $g = -d\tau^2 + g_{ij}\,dx^i\,dx^j$, because

$$d\tau = \sqrt{\tilde{r}}\,d\tilde{r} - \sqrt{r}\,dr,$$

$$dt = d\tau - dr\frac{\sqrt{r}}{r-1} = d\tau\frac{r}{r-1} - d\tilde{r}\frac{\sqrt{\tilde{r}}}{r-1}$$

leads to

$$\begin{aligned} g &= -dt^2\frac{r-1}{r} + dr^2\frac{r}{r-1} + r^2\,d\Omega^2 \\ &= -d\tau^2 + (1 - \tfrac{3}{2}\tau\tilde{r}^{-3/2})^{-2/3}\,d\tilde{r}^2 + (1 - \tfrac{3}{2}\tau\tilde{r}^{-3/2})^{4/3}\tilde{r}^2\,d\Omega^2. \end{aligned} \tag{4.5.28}$$

**Remarks** (4.5.29)

1. This chart can be used for $3\tau/2 < \tilde{r}^{3/2}$, though at $3\tau/2 = \tilde{r}^{3/2}$, which corresponds to $r = 0$, it becomes singular. Equation (4.4.28) therefore extends the Schwarzschild metric beyond $r = r_0 = 1$, but it is not the maximal extension (4.4.48).
2. A particle falling freely from infinity travels from $r = \tilde{r}$ to $r = 0$ in proper time $2\tilde{r}^{3/2}/3$, in units where $r_0 = 1$.

To discover the proper Friedmann solution, note that Equation (4.5.19), with $a = R(0)$, $r_0 = 8\pi\kappa\rho(0)a^3/3 = 1$, and $K = 0$, implies that

$$\dot{R} = -\frac{1}{\sqrt{R}},$$

and hence that

$$R(t) = a(1 - \tfrac{3}{2}ta^{-3/2})^{2/3}.$$

By redefining the coordinates $t \to \tau$, $ar \to \tilde{r}$, we come up with the

**Oppenheimer–Snyder Solution** (4.5.30)

*The metric*

$$g = -d\tau^2 + \begin{cases} (1 - \frac{3}{2}\tau a^{-3/2})^{4/3}(d\tilde{r}^2 + \tilde{r}^2\, d\Omega^2) & \text{if } \tilde{r} \le a \\ (1 - \frac{3}{2}\tau\tilde{r}^{-3/2})^{-2/3}\, d\tilde{r}^2 + (1 - \frac{3}{2}\tau\tilde{r}^{-3/2})^{4/3}\tilde{r}^2\, d\Omega^2 & \text{if } \tilde{r} \ge a, \end{cases}$$

*satisfies Einstein's equations with* $p = 0$,

$$\rho = \begin{cases} 3(a^{3/2} - 3\tau/2)^{-2}/8\pi\kappa & \text{if } \tilde{r} \le a \\ 0 & \text{if } \tilde{r} > a. \end{cases}$$

**Proof**

Einstein's equations are satisfied for $\tilde{r} > a$ and $\tilde{r} < a$ by construction, and, as the curvature ought to be discontinuous at $\tilde{r} = a$ (cf. (4.2.13)), it is only necessary to check that it has no delta-function singularity there. Let us write the orthogonal basis for $g$,

$$e^{\alpha} = (d\tau, e^{\nu}v'\, d\tilde{r}, e^{\nu}\, d\vartheta, e^{\nu} \sin\vartheta\, d\varphi)$$

$$e^{\nu(\tau,\tilde{r})} \equiv \begin{cases} \tilde{r}(1 - \frac{3}{2}\tau a^{-3/2})^{2/3} & \text{if } \tilde{r} \le a \\ \tilde{r}(1 - \frac{3}{2}\tau\tilde{r}^{-3/2})^{2/3} & \text{if } \tilde{r} \ge a, \end{cases} \tag{4.5.31}$$

where $v' \equiv \partial v/\partial\tilde{r}$ and $\dot{v} \equiv \partial v\ \partial\tau$; $v'$ is discontinuous and $\dot{v}$ is continuous with a discontinuous first derivative. When restricted to $r = a$ ($d\tilde{r}_{|r=a} = 0$), only the continuous parts of $e^{\alpha}$ remain, and even the $de^{\alpha}$ are continuous at $\tilde{r} = a$:

$$de^{\alpha} = (0, e^{\nu}(\dot{v}' + \dot{v}v')d\tau \wedge d\tilde{r}, e^{\nu}(\dot{v}\, d\tau + v'\, d\tilde{r}) \wedge d\vartheta,$$
$$e^{\nu}((\dot{v}\, d\tau + v'\, d\tilde{r})\sin\vartheta + \cos\vartheta\, d\vartheta) \wedge d\varphi).$$

From this formula we get the affine connections,

$$\omega^{\alpha}{}_{\beta} = \begin{array}{c|cccc} & \tau & \tilde{r} & \vartheta & \varphi \\ \hline \tau & 0 & e^{\nu}(\dot{v}' + \dot{v}v')d\tilde{r} & e^{\nu}\dot{v}\, d\vartheta & e^{\nu}\dot{v} \sin\vartheta\, d\varphi \\ \tilde{r} & & 0 & -d\vartheta & -\sin\vartheta\, d\varphi \\ \vartheta & & & 0 & -\cos\vartheta\, d\varphi \\ \varphi & & & & 0 \end{array} \tag{4.5.32}$$

Observe that the discontinuous functions $v'$ and $\dot{v}'$ are multiplied by $d\tilde{r}$, and so no $v''$ shows up in $d\omega^{\alpha}{}_{\beta}$ (see Problem 4). Hence, while $R^{\alpha}{}_{\beta}$ is discontinuous, it does not contain a delta function. □

**Remarks** (4.5.33)

1. As Einstein's equations do not allow delta-function singularities in the contractions $R_\alpha$, the question arises of whether they can occur in $R^\alpha{}_\beta$. The answer is that they can not, because the surface of discontinuity has a spacelike normal $d\tilde{r}$, and by (4.2.24), the regularity of $R_\alpha$ implies that of $R^\alpha{}_\beta$ in this case.
2. If the basis given in the proof is supplied with a more general function $v(\tau, r)$, then it is easy to find a solution of Einstein's equations that describes the gravitational collapse of a $C^\infty$ density distribution. The discontinuous solution (4.5.30) can be considered as the limiting case of a $C^\infty$ solution (Problem 4).

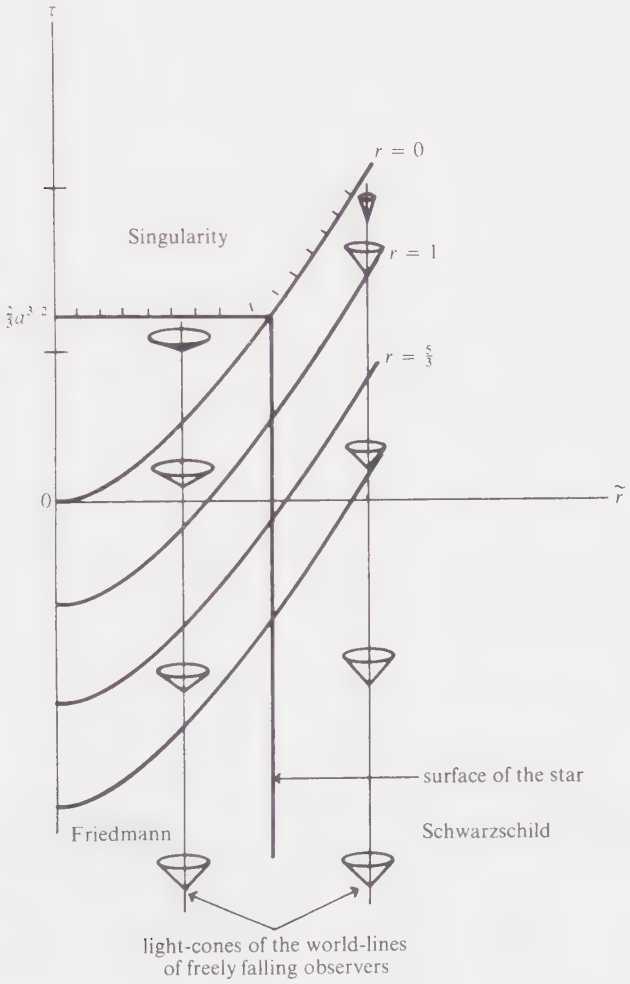



Figure 65

3. In order that the solutions join without a seam, the Schwarzschild radius of the outer solution must be

$$8\pi\kappa \int_0^a d\tilde{r}\, \tilde{r}^2 \rho(\tilde{r}, \tau)_{|\tau=0}.$$

This is to be expected, because when $\tau = 0$, $dr = d\tilde{r}$. In the static coordinates $(t, r)$, the dramatic action in the interior is not detectable from outside. The Schwarzschild solution is as silent as the grave about what it hides within.

**The Geometric Significance of the Oppenheimer–Snyder Solution** (4.5.34)

The chart of (4.5.30) is workable if

$$\tau > \tfrac{2}{3}\min[a^{3/2}, \tilde{r}^{3/2}],$$

though the metric becomes singular at the boundary. The lines $\tilde{r} =$ constant are the world-lines of freely falling observers, and $\tau$ measures their proper time. In this chart the light cones inside the star flatten out as they approach the singularity, while on the outside they narrow down (see Figure 65). Note that the significance of $r = 1$ as a horizon for $\tilde{r} > a$ can thereby be expressed as the fact that from this point on, the light cones remain completely on one side of the curve $r =$ constant.

**Problems** (4.5.35)

1. Integrate (4.5.11) for $\rho =$ constant (cf. (4.4.12)).
2. Show that $\rho = 3/56\pi\kappa r^2$ is a solution of (4.5.11) for the equation of state of radiation, $p = \rho/3$.
3. Solve (4.4.10) for $K > 0$, if $p = \rho/3$ and if $p = \rho$.
4. Use the basis (4.5.31)

$$e^{\alpha} = (d\tau, e^{\nu}\nu'\, d\tilde{r}, e^{\nu}\, d\vartheta, e^{\nu} \sin\vartheta\, d\varphi), \quad \text{with} \quad \nu(\tau, \tilde{r}),$$

to find a solution of Einstein's equations for $\mathcal{T}_0 = \rho(\tilde{r})e^0$, $\mathcal{T}_j = 0$.

**Solutions** (4.5.36)

1. Let $x$ be the dimensionless variable given in (4.5.13), and $y = p/\rho$; then this equation becomes

$$-\frac{dy}{dx} = (1+y)(1+3y)\frac{x/2}{1-x^2} \quad \text{i.e.,} \quad dy\left(\frac{-3}{1+3y} + \frac{1}{1+y}\right) = \frac{dx\, x}{1-x^2}$$

$$\Rightarrow \ln\frac{1+3y}{1+y} = \frac{1}{2}\ln\frac{1-x^2}{1-x_0^2} \Rightarrow \frac{1+3y}{1+y} = \sqrt{\frac{1-x^2}{1-x_0^2}}$$

$$\Rightarrow \frac{\sqrt{1-x^2} - \sqrt{1-x_0^2}}{3\sqrt{1-x_0^2} - \sqrt{1-x^2}} = y.$$

2. $p = 1/56\pi\kappa r^2, \qquad M(r) = 4\pi\kappa \int_0^r dr'\, r'^2\, \rho(r') = 3r/14,$

$$-\frac{\partial p}{\partial r} = \frac{1}{28\pi\kappa r^3} = \frac{4}{56\pi\kappa r^3}\,\frac{\frac{3}{14}r + r/14}{r(1-\frac{3}{7})} = \frac{\rho + p}{r}\,\frac{M(r) + 4\pi\kappa p r^3}{r - 2M(r)}.$$

3. $p = \frac{\rho}{3}$: $\quad \frac{K + \dot{R}^2}{R^2} = -\frac{2R\ddot{R} + \dot{R}^2 + K}{R^2} \Rightarrow \dot{R}^2 + R\ddot{R} + K = 0 \Rightarrow \frac{d^2}{dt^2}R^2 = -2K$

$$\Rightarrow R = (ct - Kt^2)^{1/2}.$$

$p = \rho$: Let $d\tau/dt = 1/R$. Then $4K + 4\dot{R}^2 + 2R\ddot{R} = 0$ becomes the oscillator equation,

$$\frac{d^2R^2}{d\tau^2} = -4KR^2 \Rightarrow R = R_{\max}[\sin 2\sqrt{K}\,\tau]^{1/2};$$

consequently $t$ is given by

$$R_{\max}\int_0^\tau d\tau'[\sin 2\sqrt{K}\,\tau']^{1/2}.$$

In both cases we have chosen $R(0) = 0$, and we observe that $R$ decreases again to zero after a finite time.

4. By (4.1.28) the affine connections (4.5.32) lead to the curvature forms

$$\begin{aligned}
R^0{}_{\tilde{r}} &= e^{\nu}(2\dot{\nu}\nu' + \dot{\nu}^2\nu' + \ddot{\nu}' + \ddot{\nu}\nu')d\tau \wedge d\tilde{r}\\
R^0{}_{\vartheta} &= e^{\nu}(\dot{\nu}^2 + \ddot{\nu})d\tau \wedge d\vartheta\\
R^0{}_{\varphi} &= e^{\nu}(\dot{\nu}^2 + \ddot{\nu})\sin\vartheta\, d\tau \wedge d\varphi\\
R^{\tilde{r}}{}_{\vartheta} &= e^{2\nu}(\dot{\nu}' + \dot{\nu}\nu')\dot{\nu}\, d\tilde{r} \wedge d\vartheta\\
R^{\tilde{r}}{}_{\varphi} &= e^{2\nu}(\dot{\nu}' + \dot{\nu}\nu')\dot{\nu}\sin\vartheta\, d\tilde{r} \wedge d\varphi\\
R^{\vartheta}{}_{\varphi} &= e^{2\nu}\dot{\nu}^2\sin\vartheta\, d\vartheta \wedge d\varphi.
\end{aligned}$$

Once again, $R_{\alpha\beta} \sim e_{\alpha\beta}$, so $T_{\alpha\beta}$ is diagonal. Einstein's equations require that

$$\begin{aligned}
8\pi\kappa\rho &= 3\dot{\nu}^2 + 2\frac{\dot{\nu}'\dot{\nu}}{\nu'} = 8\pi\kappa T_{00}\\
0 &= e^{\nu}(3\dot{\nu}^2 + 2\ddot{\nu}) = T_{\tilde{r}\tilde{r}}\\
0 &= \left(1 + \frac{1}{2\nu'}\frac{d}{d\tilde{r}}\right)(3\dot{\nu}^2 + 2\ddot{\nu}) = T_{\vartheta\vartheta} = T_{\varphi\varphi}.
\end{aligned}$$

The last two of these equations are solved by $e^{\nu} = (F(\tilde{r})\tau + G(\tilde{r}))^{2/3}$. Since the basis is invariant under a change of charts $\tilde{r} \to \bar{r}(\tilde{r})$, we may set $G = \tilde{r}^{3/2}$, and are left with only one function, $F(\tilde{r})$. If $\tau = 0$, then the first of Einstein's equations becomes

$$FF' = 9\pi\kappa\tilde{r}^2\rho(\tilde{r}, 0) \Rightarrow F(\tilde{r}) = -\left[18\pi\kappa\int_0^{\tilde{r}} dr\, r^2\rho(r, 0)\right]^{1/2}$$

In the case $\rho(r, 0) = 3/8\pi\kappa a^3$, we revert to (4.5.30).

## 4.6 The Existence of Singularities

*The solutions of nonlinear differential equations have a tendency to develop singularities; in particular this is true of Einstein's equations, where the attractive nature of gravity reveals its physical origins.*

The models we have considered of gravitational collapse, in which a singularity develops, were all radially symmetric. It is not at all surprising that a fall directed right at the center will end in a catastrophe. The one new feature of Einstein's theory is that the catastrophe can not be prevented by any pressure, no matter how strong, because the pressure itself produces more gravity. There is a question, however, whether the situation is qualitatively changed by a perturbation of the radial symmetry, just as the angular momentum in the Kepler problem prevents the plunge into the center. In the relativistic Kepler problem, the effective gravitational potential goes as $-1/r^3$, which is stronger than the centrifugal potential (cf. (I: §5.7)), but it is conceivable that other mechanisms might impede the growth of a singularity. It is often claimed ([32], §1.14), on account of this observation, that normally the solutions are free of singularities, which are pathologies afflicting spaces of high symmetry. It was the accomplishment of R. Penrose and others of the school of D. Sciama to disprove this claim: as long as energy and pressure are positive in some reasonable sense, and at some instant there exists the kind of geometry set up by a large mass, then the formation of a singularity is unavoidable, regardless of any symmetry.

Let us agree at this stage what we mean by a singular space. Regularity is incorporated in the concept of a manifold, and any singular points are removed. It might be suggested that unbounded growth of $R^{\alpha}{}_{\beta}$ could be taken as a sign of a singularity in the vicinity. We shall see shortly that the $R^{\alpha}{}_{\beta}$ describe the tidal force, and consequently this conjectured indicium has a direct physical significance, as it can be observed as bodily discomfort. Unfortunately, it is difficult to express this mathematically, since the components of $R^{\alpha}{}_{\beta}$ depend on the basis, and could also become infinite in the absence of a genuine singularity. Conversely, it is possible for all 14 of the invariants that can be constructed from $R^{\alpha}{}_{\beta}$ to vanish without $R^{\alpha}{}_{\beta}$ itself vanishing. For example, this happens for plane gravitational waves, and is analogous to a nonzero vector in Minkowski space having zero length.

Hence we resort to a different feature of the solutions we have discussed as the criterion for a singularity, viz., that an observer falls into the singularity in a finite proper time, thus leaving the manifold. There is, of course, the trivial possibility that the manifold has simply been chosen too small—if the manifold were only a piece of Minkowski space, then one could leave it in a finite time, although there is not necessarily any singularity outside the piece. In order to exclude such cases, we make a

**Definition** (4.6.1)

A pseudo-Riemannian manifold $M$ is **extensible** iff it is a proper subset of a larger manifold $M'$, i.e., its metric is the restriction to $M$ of the metric on $M'$.

**Remarks** (4.6.2)

1. $M'$ is not, of course, uniquely determined by $M$, so our criterion can not involve examining an extensible manifold to see where there are singularities. For instance, the Schwarzschild metric for $r > 5r_0$ can be extended either to the regular solution with a continuous mass distribution for $r < 5r_0$ or to the singular solution.
2. When confronted with an extensible manifold, one gets the feeling that something has been intentionally left out. Therefore we postulate that the physical space-time continuum is nonextensible.
3. There are examples ([33], p. 58) of nonextensible manifolds that can be escaped from, so it is actually necessary to postulate a more refined property, local nonextensibility. However, the examples seem rather artificial, so we shall content ourselves with the primitive definition.

The next step is to decide what observers we will grant an unlimited stay in the manifold.

**Definition** (4.6.3)

A pseudo-Riemannian manifold is said to be **geodesically complete** in timelike directions iff every timelike geodesic can be extended to an arbitrarily long proper-time parameter.

**Remarks** (4.6.4)

1. A positive metric $g$ defines a metric for the topology of a Riemannian space $M$, and then geodesic completeness means the same thing as completeness in the sense of a metric topological space.
2. An affine parameter could also be defined on lightlike geodesic lines, and one can speak of lightlike and spacelike geodesic completeness. These conditions are not equivalent (Problem 1); but at any rate (4.6.3) must be required on physical grounds.
3. Geodesic incompleteness puts an observer who can stay in the manifold for only a finite time into a predicament, but is not necessarily evidence

of any kind of infinity. This is shown by the example

$$g = -dt^2\left(1 - \frac{h}{2}\right) + dx^2\left(1 + \frac{h}{2}\right) + dx\,dt\,h,$$

$$h = \frac{\lambda}{2}(\cos^4(t - x) - 1), \qquad \lambda \in (0, 2),$$

on $\mathbb{R}^2$ (if desired, $dy^2 + dz^2$ can be added in). If $\lambda$ is small, this is only a weak gravitational wave that spreads throughout the flat space, but nonetheless the space fails to be geodesically complete in timelike directions, even for arbitrarily small $\lambda$ (Problem 2). The reason is that a particle of the right initial velocity rides the crests of the waves, as in a linear accelerator, and reaches nearly the speed of light. Its proper time runs ever more slowly and never exceeds some finite value. There is no singularity, and the only $R^t{}_j$ that does not vanish is $R^0{}_1{}^0{}_1 = e^0{}_1 h'' 2$. Since $h$ is periodic in $u = t - x$, $g$ can be used as a pseudometric on $T^2$, in which case even this compact set is geodesically incomplete, although it is certainly not a piece of a larger connected manifold. Despite that, we follow common usage and refer to the space as singular.

4. Even in Minkowski space it is possible to reach the end of the manifold after a finite proper time on certain timelike lines. If, for instance, $x = t + 1/t^2$ for $t > 1$, then

$$\int ds = \int \frac{ds}{dt}\,dt = \int_1^\infty dt \sqrt{1 - \left(\frac{dx}{dt}\right)^2} < 2\int_1^\infty dt\, t^{-3/2} = 4.$$

It is only the choice of coordinates that makes the end at $x = \infty$, and it can be transformed to any finite point, just as the end lay at $-\infty$ in the Schwarzschild metric with the variable $\ln r$.

5. One might require that timelike lines with bounded acceleration $\ddot{z}^\alpha \ddot{z}^\beta g_{\alpha\beta}$ can be continued to arbitrarily long proper times. If this were not so, then the crew of a rocket with a finite supply of fuel could conceivably find themselves at the edge of the universe, and would not know what to do. Yet geodesic completeness leaves this possibility open [34].

Geodesic lines are the world-lines of freely falling observers (cf. Problem 4). The nonrelativistic analogue of a geodesic vector field is the velocity field $v_i$ of an ideal fluid with no pressure in a gravitational potential $\Phi$. For stationary fluid flow, the equations of hydrodynamics require that $v_k v_{i,k} = -\Phi_{,i}$. Let $n$ be the distance-vector field between nearby fluid particles, which is carried along with the stream. Its Lie derivative with respect to $v$ vanishes, so

$$v_k n_{i,k} = n_k v_{i,k} \tag{4.6.5}$$

(recall (1: 2.5.12; 5)), which makes the second derivative along the streamlines

$$v_k \frac{\partial}{\partial x^k}\left(v_j \frac{\partial}{\partial x^j} n_i\right) = -n_k \Phi_{,ik}. \tag{4.6.6}$$

Thus the gradient of the field $\Phi_{,i}$ affects the distance between two particles, and in fact the effect of the second derivative of $\Phi$ is to focus them together: Since $\Phi$ satisfies the equation

$$\Phi_{,jj} = \rho \geq 0, \tag{4.6.7}$$

the net effect of the gravitational field, when averaged over all directions, is to focus particles. For irrotational fluid flow, $v_{i,k} = v_{k,i}$, this can be expressed as an increase in the rate of convergence $c = -v_{i,i}$ of the flow along the streamlines:

$$v_i \frac{\partial}{\partial x_i} c = -v_i v_{k,ik} = v_{k,i} v_{i,k} + \Phi_{,kk} \geq \frac{c^2}{3}. \tag{4.6.8}$$

This equation used (4.6.7) and the irrotationality, which entered through the trace inequality for symmetric $n \times n$ matrices

$$(\operatorname{Tr} M)^2 \leq n \operatorname{Tr}(M^2) \tag{4.6.9}$$

(Problem 3). If $c$ is positive at some point, then it increases so rapidly by (4.6.8) that it soon reaches infinity, and the streamlines meet. If $s$ is the parameter on a streamline, given as $x(s)$, $v_i(x) = dx_i/ds$ (cf. (4.1.42; 1)), then (4.6.8) implies that

$$\frac{dc}{ds} \geq \frac{c^2}{3} \Rightarrow c(s) \geq \frac{c(0)}{1 - sc(0)/3}, \tag{4.6.10}$$

and thus $c$ gets arbitrarily large before $s = 3/c(0)$. This elementary property of gravity contains the essential features of the relativistic theory discussed below.

The relativistic generalizations of (4.6.6) are

**The Equations of Geodesic Deviation** (4.6.11)

*Let $v = v^\alpha e_\alpha$ be a geodesic vector field and $n$ a vector field such that $L_v n = 0$. Then*

$$D_v D_v n = -e_\alpha (R^\alpha{}_\beta | n \otimes v) v^\beta.$$

**Proof**

By (4.1.7(g')), $D_v n = D_n v$, and because of (4.1.19), (4.1.33; 2), and the equation $D_v v = 0$,

$$\begin{aligned} 0 = D_n D_v v &= D_v D_v n + (D_n D_v - D_v D_n) v \\ &= D_v D_v n + e_\alpha (R^\alpha{}_\beta | n \otimes v) v^\beta. \end{aligned}$$

□

**Example** (4.6.12)

Consider the Friedmann universe (4.4.2). Let the fields $v$ and $n$ be the natural contravariant bases $\partial_t$ and $\partial_x$. Their Lie brackets with each other vanish, and $\partial_t$ is geodesic, since we are using co-falling coordinates. The contravariant components of the metric (4.4.2) are

$$g^{00} = -1, \qquad g^{jj} = \left(\frac{1 + Kr^2/4}{R}\right)^2,$$

so $v$ and $n$ can be written in the orthogonal basis as

$$v = \partial_0 = e_0, \qquad n = \partial_1 = \frac{e_1}{\sqrt{g^{11}}} = e_1 \frac{R}{1 + Kr^2/4}.$$

With the affine connections of (4.4.5), we see that

$$D_v n = \frac{\dot{R}}{1 + Kr^2/4} e_1 + \frac{R}{1 + Kr^2/4} D_v e_1 = \frac{\dot{R}}{1 + Kr^2/4} e_1,$$

$$D_v D_v n = \left(D_v \frac{\dot{R}}{1 + Kr^2/4}\right) e_1 = \frac{\ddot{R}}{1 + Kr^2/4} e_1 = \frac{\ddot{R}}{R} n,$$

which, because of (4.4.6), is precisely

$$-e_1(R^1{}_0 | n \otimes v).$$

**Remarks** (4.6.13)

1. Since $R(t)$ describes how the distance between neighboring world-lines $x =$ constant varies, we perceive that $D_t D_t n$ has the significance of a relative acceleration.
2. Proposition (4.6.11) shows that from the physical point of view it is $R^{\alpha}{}_{\beta}$ rather than $\omega^{\alpha}{}_{\beta}$ that takes over the role of the electric field strength. Because of the principle of equivalence (I: 5.6.11), there is no trace of the $\omega$'s; freely falling observers can only notice the gradient of the field $R$, specifying the corrections to the principle of equivalence, which holds only in the infinitesimal limit.

The curvature forms may have either sign, either focusing or defocusing. The contractions $R_{\alpha}$ are immediately determined by the energy and momentum, from which they inherit the positivity (2.1.13). As with (4.6.8), this leads to an

**Increase in the Rate of Convergence of Geodesic Vector Fields** (4.6.14)

*Let $v$ be a timelike geodesic vector field perpendicular to a hyperplane $t = 0$, and assume $i_v R_0 \geq 0$. Then the rate of convergence $c = -\sum_{\alpha} \langle e^{\alpha} | D_{e_{\alpha}} v \rangle$ (cf. Problem 4) satisfies the differential inequality*

$$D_v c \geq \frac{c^2}{3}.$$

**Proof**

We work in the natural basis of a co-falling coordinate system, so that $v = \partial_t$ and $g = -dt^2 + g_{ab}\, dx^a \otimes dx^b$. As in (4.6.11), $D_v v = 0$, and $D_v \partial_a = D_{\partial_a} v$. By also recalling that

$$0 = D_v \delta^\alpha{}_\beta = \langle D_v\, dx^\alpha | \partial_\beta \rangle + \langle dx^\alpha | D_v\, \partial_\beta \rangle,$$

we find

$$\begin{aligned} D_v c &= -D_v \langle dx^\alpha | D_{\partial_\alpha} v \rangle = -\langle D_v\, dx^\alpha | D_v\, \partial_\alpha \rangle - \langle dx^\alpha | D_v D_{\partial_\alpha} v \rangle \\ &= -\langle D_v\, dx^\alpha | \partial_\beta \rangle \langle dx^\beta | D_v\, \partial_\alpha \rangle + \langle R^\alpha{}_\beta | \partial_\alpha \otimes v \rangle v^\beta \\ &= \langle dx^\alpha | D_v\, \partial_\beta \rangle \langle dx^\beta | D_v\, \partial_\alpha \rangle + i_v R_0 \geq \frac{c^2}{3}, \end{aligned}$$

since the trace inequality (4.6.9) is again applicable:

$$M_{\alpha\beta} \equiv \langle \partial_\alpha | D_v\, \partial_\beta \rangle = \langle \partial_\alpha | D_{\partial_\beta} v \rangle = \Gamma^0_{\alpha\beta}$$

is symmetric in $\alpha$ and $\beta$ and vanishes when $\alpha$ or $\beta$ is zero, because of $D_v v = 0$. In the space orthogonal to $v$, $g$ is positive and $c = \mathrm{Tr}(Mg) = \mathrm{Tr}(\sqrt{g} M \sqrt{g})$, while $\mathrm{Tr}(\sqrt{g} M \sqrt{g} \sqrt{g} M \sqrt{g})$ occurs in the above equation. □

**Example** (4.6.15)

Let us take another look at the Friedman universe, for which

$$D_{\partial_a} v = D_v\, \partial_a = \frac{\dot{R}}{R} \partial_a,$$

and thus $c = -3\dot{R}/R$. Proposition (4.6.14):

$$D_v c = -\frac{\partial}{\partial t} \frac{3\dot{R}}{R} = \frac{3\dot{R}^2}{R^2} - \frac{3\ddot{R}}{R} \geq \frac{c^2}{3} = \frac{3\dot{R}^2}{R^2}$$

holds if $\ddot{R} \leq 0$, which amounts to the condition that $i_v R_0 \geq 0$. From Equations (4.4.10), this condition is met if $\rho + 3p \geq 0$.

**Remarks** (4.6.16)

1. According to Einstein's equations (4.2.20(a)), the condition $i_v R_0 \geq 0$ $\forall v$: $\langle v | v \rangle < 0$ implies for the total energy-momentum tensor that $T_{00} + T_{11} + T_{22} + T_{33} \geq 0$ ("positivity of the energy"). This is true for all sensible models of matter. The reason for the positivity condition is that negative energy produces a repulsive gravitational force, which could prevent the convergence of the geodesics.

2. Generalizing the nonrelativistic result (4.6.10), we see that if $c$ is ever positive, then it must become infinite after a finite time under the circumstances of (4.6.14). Let $s$ be the proper-time parameter (cf. (4.1.42; 1)) on the geodesics of $v$, $v^i(x(s)) \equiv \dot{x}^i(s)$, and $c(s) \equiv c(x(s))$. If $c(0) > 0$, then $c$ becomes infinite for some $s$ such that $0 \leq s \leq 3/c(0)$.
3. If $N_{ab} = \langle e_a | e_b \rangle$, then $(\mathrm{Det}\, N)^{1/2}$ is the volume spanned by the spacelike basis vectors. Since

$$c = -\langle e^a | D_v e_a \rangle = -\tfrac{1}{2} D_v \ln(\mathrm{Det}\, N),$$

$c \to \infty$ would mean that the volume approaches zero, and neighboring geodesic lines meet. Hence there exists a basis field $n$, $L_v n = 0$, which becomes zero.

It is not necessarily a sign of a disaster if $c$ does not remain bounded; (4.6.14) made no use of the strict positivity of $i, R_0$, which means that the same conclusions could be reached in flat space, even though the metric is everywhere regular.

**Example** (4.6.17)

Let us introduce the coordinates $\tau = -\sqrt{t^2 - x^2}$ and $u = x/t$ on $M = \{(t, x) \in \mathbb{R}^2 : t < 0, x < 0, t^2 > x^2; g = -dt^2 + dx^2\}$, and let the vector fields

$$v = \partial_\tau = -\frac{t}{\sqrt{t^2 - x^2}}\,\partial_t - \frac{x}{\sqrt{t^2 - x^2}}\,\partial_x,$$

$$n = \partial_u = \frac{xt^2}{t^2 - x^2}\,\partial_t + \frac{t^3}{t^2 - x^2}\,\partial x.$$

serve as the basis; it is easy to verify that $D_v v = 0, D_v \tau = 1$, and $D_v n = D_n v = -n/\tau$. The streamlines of $v$ are geodesic since they are straight lines through the origin, and the streamlines of $n$ are normal to them in the sense of the metric (Figure 66).

$$c = -\langle v | D_v v \rangle - \frac{\langle n, D_v n \rangle}{\langle n | n \rangle} = \frac{1}{\tau},$$

and in fact

$$D_v c = -\frac{\partial}{\partial \tau}\frac{1}{\tau} = \frac{1}{\tau^2} = c^2,$$

and in one space dimension (4.6.14) can be strengthened to $D_v c \geq c^2$. At the origin $c$ becomes infinite, although the space is not singular; it is only that the chart $(\tau, u)$ is unsuitable there.

When geodesics cross, they lose the property of being extremal, which can bring about some contradictions. This will lead to the conclusion that one

Figure 66 The convergence of a geodesic vector field on flat space

can not make the assumption that geodesics are extensible past the points where they cross. Let us distinguish the different possible states of affairs by making

**Definition** (4.6.18)

(a) The **future** $J^+(x)$ (respectively **past** $J^-(x)$) of $x \in M$ consists of the points of $M$ that can be connected to $x$ by causal curves (see (2.1.15)) directed toward the past (respectively future).
(b) Let $S$ be a spacelike hypersurface and $x \in D^+(S)$. Then the set of causal (respectively, differentiable causal) curves from $x$ to $S$ is denoted by $C(x, S)$ (resp., $C^1(x, S)$). The set $C(x, S)$ is topologized as follows: A basis of neighborhoods of the curve $\lambda$ consists of all curves that stay in a neighborhood of $\lambda$. in the sense of the topology of $M$. We let $C^1$ have the topology induced by $C$.
(c) The **length**, or **proper time**, of $\lambda \in C^1(x, S)$ is defined by

$$d(\lambda) = \int_{s_0}^{s_1} ds \sqrt{-g_{\alpha\beta}\dot{\lambda}^\alpha \dot{\lambda}^\beta}, \qquad \lambda(s_0) \in S, \qquad \lambda(s_1) = x.$$

**Remarks** (4.6.19)

1. The topology on $C(x, S)$ is that of uniform convergence. It is metrizable, since the topology of $M$ is metrizable: the Hausdorff distance function between two subsets of $M$ is defined in terms of the topology of $M$, and the distance function of two causal curves produces the metric on $C(x, S)$. Consequently, compactness becomes synonymous with sequential compactness.
2. $C^1$ is dense in $C$, and so $d(\lambda)$ can be extended to $C$ (Problem 5).

To simplify the next chain of reasoning, let us eliminate pathologies at the outset by making for the duration the

**Assumptions** (4.6.20)

(a) that $M$ is orientable with respect to time (2.1.15); and
(b) that if $x$ and $y$ vary over $M$, then the interiors of $J^-(x) \cap J^+(x)$ form a basis for the topology of $M$ (cf. (I: 6.4.10; 3)).

The above assumptions imply the

**Propositions** (4.6.21)

*For all $x$ in the interior of $D^+(S)$,*
(a) *$J^-(x) \cap D^+(S)$ is compact;*
(b) *$C(x, S)$ is compact; and*
(c) *$d: C^1(x, S) \to \mathbb{R}^+$ is upper semicontinuous.*

**Remarks** (4.6.22)

1. Neither time direction is distinguished, either here or below. Where appropriate, $D^-$ (and $J^-$) can be substituted for $D^+$ (and $J^+$).
2. Proposition (b) is a variant of Ascoli's theorem, according to which any family of equicontinuous curves on compact sets is relatively compact in the topology of uniform convergence. However, a set of curves of arbitrary gradient is not compact; for example, $x = \sin nt$, $n = 1, 2, \ldots$ is not uniformly convergent to anything. The requirement that a curve never gets out of a light-cone prevents this from happening in (4.6.21).
3. Proposition (a) is a necessary condition for (b); for example, $x = \sin t/n$, $n = 1, 2, \ldots$, does not converge uniformly to zero on $-\infty < t \leq 0$.
4. The function $d$ is not continuous, because in any neighborhood of $\lambda \in C(x, S)$ it is possible to reflect lightlike curves back and forth to make $d$ vanish.
5. The extension of $d$ to $C(x, S)$ is likewise upper semi-continuous (Problem 5). We let it define the proper time on a nondifferentiable causal curve.

The proofs of these propositions are rather technical, and are left for Problem 6. An important consequence of them is

**Theorem** (4.6.23)

*Let $S$ be a spacelike hypersurface and $p$ be in the interior of $D^+(S)$. Then there is a curve of greatest proper time from $p$ to $S$, and it is the geodesic through $p$ that is orthogonal to $S$ in the sense of the pseudometric $g$.*

**Proof**

It follows from the compactness of $C$ and the upper semicontinuity of $d$ that $d$ achieves its supremum ([22], 12.7.9). The maximal curve must be geodesic, because otherwise one could find a nearby curve of greater proper time to the same point of intersection with $S$. The orthogonality follows from the requirement that geodesics to other nearby points of $S$ take less time: according to (I: 3.2.18; 6), the change in the time taken on a geodesic line having endpoint $x$ is

$$\delta x^\alpha \frac{\partial}{\partial \dot{x}^\alpha} \sqrt{-\dot{x}^\beta \dot{x}^\gamma g_{\beta\gamma}} = \frac{\delta x^\alpha \dot{x}^\beta g_{\alpha\beta}}{\sqrt{-\dot{x}^\beta \dot{x}^\gamma g_{\beta\gamma}}}.$$

If $\dot{x}$ were not perpendicular to all the tangent vectors of $S$, then one could find a way to increase the time. □

**Example** (4.6.24)

As in Example (4.6.17), let $g = -dt^2 + dx^2$, $p = (t_0, 0)$, and $S = \{(t, x) \in \mathbb{R}^2 : t^2 - x^2 = 1, t \leq -1\}$. The straight lines through the origin are the geodesics perpendicular to $S$. The straight lines through $p$, $x = v(t - t_0)$, intersect $S$ where $x^2 = v^2(t - t_0)^2 = t^2 - 1$, so the distance along them to $S$ is $(t_0 - t)\sqrt{1 - v^2} = (1 + t_0^2 - 2t_0 t)^{1/2}$ (see Figure 67). If $t_0 \leq 0$, the maximum is achieved at $t = -1$, since $t \leq -1$, and the geodesic orthogonal to $S$ is then the longest line. If $t_0 > 0$, then the distance grows without bound

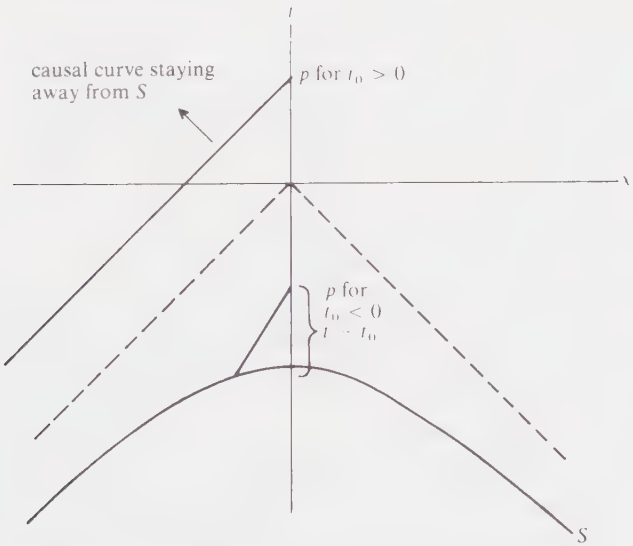



Figure 67 Geodesic lines perpendicular to $S$ in (4.6.24)

as $t \to \infty$; there exists no maximum, and the line through $(-1, 0)$ is the shortest. This does not contradict (4.6.23), since in this case $p \notin D^+(S)$.

It is intuitively reasonable that one could obtain a more nearly extremal curve from two crossing geodesics, by rounding them off near the intersection. This expectation is confirmed by the

**Theorem** (4.6.25)

*Let $v$: $D_v v = 0$ and $\langle v|v\rangle = -1$ be the geodesic vector field perpendicular to a hypersurface $S$, let $\gamma(\tau)$ be a streamline of $v$ and suppose $n$ with the properties $L_v n = 0$ and $\langle n|v\rangle = 0$ vanishes at $\gamma(0)$ but not on all of $\gamma$. Then for all $p > 0$, $\gamma$ is not the curve of greatest proper time from $\gamma(p)$ to $S$.*

**Proof**

We choose co-moving coordinates $g = -d\tau^2 + g_{ik}\,dx^i\,dx^k$, and let $\tau$ be the proper time along $-v = \partial_\tau$. Since $n$ satisfies (4.6.11) and vanishes at $\tau = 0$, it is impossible for $D_v n$ to vanish at that point, as it would otherwise be zero on $\gamma$. Therefore, letting $w \equiv n/\tau$, $\lim_{\tau \to 0}\langle w|w\rangle$ is positive (i.e., $n$ is spacelike). Because

$$D_w v = \frac{1}{\tau} D_n v = \frac{1}{\tau} D_v \tau w = D_v w + \frac{w}{\tau}$$

the quantity

$$\langle D_w v|w\rangle = \frac{1}{\tau}\langle w|w\rangle + \langle D_v w|w\rangle$$

approaches $+\infty$ on $\gamma$ as $\tau \to 0$. This expression has the significance of the second derivative of $\tau$ along the geodesic line in the direction of $w$: Let $\bar{w}$ be the geodesic vector field having the same direction as $w$ on $\gamma$, so $D_{\bar{w}}\bar{w} = 0$ and $w_{|\gamma} = \bar{w}_{|\gamma}$; then

$$\langle D_w v|w\rangle_{|\gamma} = D_{\bar{w}}\langle d\tau|\bar{w}\rangle_{|\gamma} = \frac{\partial^2}{\partial r^2}\tau,$$

where $r$ is the curve parameter of the streamline $dx^\alpha/dr = \bar{w}^\alpha(x(r))$. If $p - \bar{\tau}(x)$ is the distance from the point $x$ to $\gamma(p)$ along the geodesic connecting these two points, then the function $\bar{\tau}(x)$ is defined and regular in a neighborhood of $\gamma(0)$ for sufficiently small $p$. On $\gamma$, $\tau = \bar{\tau}$, and there exist points on $\gamma$ such that

$$\frac{\partial^2}{\partial r^2}\tau > \frac{\partial^2}{\partial r^2}\bar{\tau};$$

therefore there exists a point $q$ in the neighborhood of $\gamma(0)$, for which $\tau_1 > \bar{\tau}_1$. If $\gamma$ intersects $S$ at $\tau = \tau_0 < 0$, then the distance from $\gamma(p)$ to $S$ along $\gamma$ equals

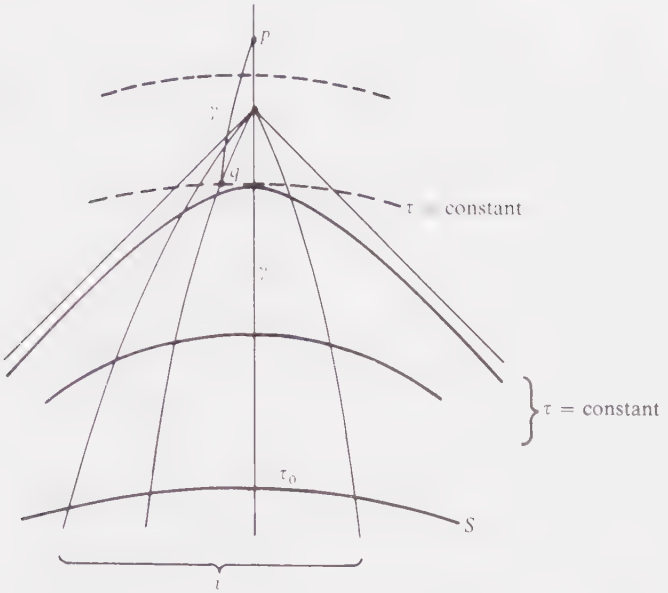



Figure 68 A curve $\gamma'$ having longer proper time than the geodesic $\gamma$

$-\tau_0 + p$. The distance along the geodesics from $p$ to $q$ is $p - \bar{\tau}_1$, and the distance from there to $S$ is $\tau_1 - \tau_0$, so that the total proper time along this path $\gamma'$ is $p - \bar{\tau}_1 + \tau_1 - \tau_0 > p - \tau_0$ (Figure 68). □

**Example** (4.6.26)

In the case of straight lines (4.6.17) and (4.6.24) with $S = \{(t, x) \in \mathbb{R}^2 : t^2 - x^2 = 1,\ t \leq -1\}$, $\gamma =$ the $t$-axis, $\tau = \sqrt{t^2 - x^2}$, and $p = (t_0, 0)$ for $t_0 > 0$, we have $\bar{w} = \partial_x$, and thus $r = x$ and

$$\frac{\partial^2}{\partial r^2}\tau_{|\gamma} = \frac{1}{|t|} > \frac{\partial^2}{\partial r^2}\bar{\tau} = \frac{\partial^2}{\partial x^2}\sqrt{(t - t_0)^2 - x^2}_{|x=0} = \frac{1}{|t - t_0|}$$

for all $t < 0$. The explicit calculation of (4.6.24) confirms the conclusions reached earlier.

Finally, let us collect our results in a

**Theorem** (4.6.27)

*Let $(M, g)$ be orientable with respect to time, $i_v R_0 \geq 0$ for timelike vectors $v$, and let $S \subset M$ be a spacelike hypersurface on which the rate of convergence of the orthogonal geodesic vector field $v$ is always $\geq c_0 > 0$. Then there can not exist a point $p$ in $D(S)$ at a distance greater than $3/c_0$ from $S$.*

**Proof**

Proposition (4.6.14) states that the rate of convergence of $r$ becomes infinite within a distance $\leq 3/c_0$ of $S$. Hence, given any geodesic through $p$ perpendicular to $S$, there exists a field $n$ that vanishes before $p$ (cf. (4.6.16; 3)), and thus no geodesic through $p$ is the curve of greatest proper time from $S$. This contradicts (4.6.23), and the only remaining logical possibility is that the geodesics can not be extended to $p$. □

**Remarks** (4.6.28)

1. If $D(S)$ is sufficiently large, for instance all of $M$ if $S$ is Cauchy, and $c$ is positively bounded on $S$, then $M$ can not be geodesically complete: there must be a singularity somewhere. The shortcoming of this statement is that we do not learn what happens physically at the singularity — whether there are infinite tidal forces, or a "quasiregular singularity" for which everything remains finite, as in (4.6.4; 3).
2. The convergence of $r$ leads one to suppose that the rate of convergence of the streamlines of matter also becomes infinite, and that the energy density is divergent at some point. The difficulty in proving this is that there might exist an earlier, quasiregular singularity, and the time-evolution might stop before reaching an infinite density.
3. In order to draw conclusions from (4.6.23) about the existence of singularities, it is necessary to know something about the size of $D(S)$. If there were a Cauchy surface $S$ with $c_0 > 0$, then singularities would be unavoidable.
4. There are numerous variations and refinements of this theorem [33], yet the precise physical nature of what happens at the singularity is still unclear.

**Examples** (4.6.29)

1. The Friedmann universe (4.6.15) with $S$ equal to the hypersurface at $t =$ constant. This is a Cauchy surface, so every later point lies in $D(S)$. Hence, if $c = -3\dot{R}/R > 0$, then no pressure, however great, can prevent the formation of a singularity.
2. Let $(M, g)$ be $(\mathbb{R}^2, -dt^2 + dx^2)$ and $S = \{(t, x): t^2 - x^2 = 1, t \leq -1, |x| \leq r\}$. Then it can be calculated that

$$D(S) = \{(t, x): t^2 - x^2 > 1, t < x + r - \sqrt{r^2 + 1}, \\ t < -x + r - \sqrt{r^2 + 1}\}$$

(see Figure 69). In this case, $c_0 = 1$ (cf. (4.6.17)), and the geodesics perpendicular to $S$ leave $D(S)$ at the latest at $r + 1 - \sqrt{1 + r^2} < 1 < 3/c_0$, so (4.6.27) predicts no singularity.

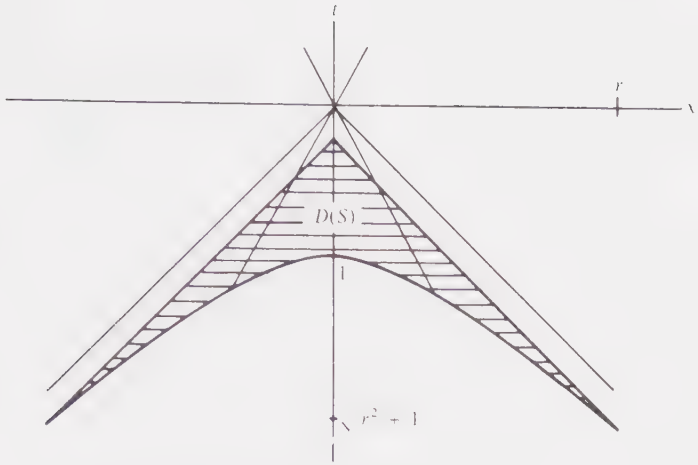



Figure 69 Intersection of geodesics in flat space

3. The Oppenheimer-Snyder Solution (4.5.30). With $S$ such that $\tau = 0$ and $v = \partial_0$, we calculate that

$$c = {}^*di_v{}^*\mathbf{1} = -\left(3\dot{v} + \frac{\dot{v}'}{v'}\right) = \begin{cases} \dfrac{3}{a^{3/2} - 3\tau/2} & \text{if } \tilde{r} < a \\ \dfrac{3}{\tilde{r}^{3/2} - 3\tau/2} & \text{if } \tilde{r} > a. \end{cases}$$

Therefore $c$ is again infinite at the singularity at $r^{3/2} = \tilde{r}^{3/2} - 3\tau/2 = 0$. In the static coordinates, $c \sim r^{-3/2}$, because the radial speed

$$v_r \sim |\text{potential energy}|^{1/2} \sim r^{-1/2}$$

means that $v_{r,r} \sim r^{-3/2}$.

The great interest in singularities is clearly because what is at stake is whether most stars end up as black holes or black holes exist only in peculiar circumstances, and whether the universe originated in a point and will some day return to a point. The question of singularities reveals how incomplete our understanding of natural laws is — are they ever superseded by something beyond human understanding?

**Problems** (4.6.30)

1. Construct an example of a pseudo-Riemannian manifold that is geodesically complete in spacelike and lightlike directions, but not in timelike directions. Do this by choosing $g = \Omega(t, x)(dx^2 - dt^2)$ on $\mathbb{R}^2$, with a suitable function $\Omega$.

2. Find timelike geodesics for the metric of (4.6.4; 3), on which $x + t$ becomes infinite after a finite proper time.

3. (a) Prove (4.6.9). (b) Find an example of a nonsymmetric, real $2 \times 2$ matrix that violates (4.6.9).

4. In the notation of (4.3.1), for the energy-momentum tensor (3.1.25; 3) with $p = 0$, $\mathcal{T}^{\alpha} = \rho v v^{\alpha}$, the contracted Bianchi identity implies that $(\rho v^{\beta} v^{\alpha})_{;\beta} = 0$. Show that this makes $v$ geodesic. Identity the rate of convergence of an arbitrary vector field as $c = -v^{\alpha}{}_{;\alpha} = -\delta v = {}^*(L_v {}^*\mathbf{1}) = {}^*(d {}^*v)$.

5. Extend the definition of $d$ from the dense set $C^1(p, S)$ to $C(p, S)$ upper semicontinuously. Give an example of a densely defined continuous function which can not be extended upper semicontinuously, and several examples of upper semicontinuous extensions of continuous functions.

6. Prove (4.6.21).

## Solutions (4.6.31)

1. Let $\Omega = 1$ for $|x| \geq 1$, $\Omega_{,x}(t, 0) = 0$, and $\lim_{t \to \infty} |t|^{2+\varepsilon} \Omega(t, 0) = 0$ for some $\varepsilon > 0$. Then the time axis is geodesic, and the proper time on it is

$$\int_{-\infty}^{\infty} ds = \int_{-\infty}^{\infty} dt \sqrt{\Omega(t, 0)} < \infty.$$

However, light rays and spacelike lines leave the strip $|x| < 1$ and continue on as in Minkowski space.

2. Let $u = t - x$ and $v = t + x$; then the Lagrangian for the motion becomes

$$\mathcal{L} \equiv \frac{h}{2} \dot{v}^2 - \dot{u}\dot{v}.$$

Consequently

$$\frac{h}{2} \dot{v}^2 - \dot{u}\dot{v} = -1 \quad \text{and} \quad P \equiv \frac{\partial \mathcal{L}}{\partial \dot{v}} = h\dot{v} - \dot{u} = \text{constant}.$$

Therefore we must integrate

$$\dot{u} = \sqrt{P^2 + 2h}, \qquad \dot{v} = \frac{1}{h}(P + \dot{u}) = \frac{2}{\dot{u} - P}.$$

If $P^2 \equiv \lambda$, then

$$\dot{u} = P \cos^2 u \Rightarrow Ps = \tan u \Rightarrow \dot{u} = \frac{P}{P^2 s^2 + 1}$$

$$\Rightarrow \dot{v} = -2 \frac{1 + 1/P^2 s^2}{P} \Rightarrow v = -\frac{2s}{P} + \frac{2}{sP^3} + \text{constant}.$$

3. (a) If $M = TmT^{-1}$, where $m$ is diagonal, with eigenvalues $m_i$, then (4.6.9) is the Cauchy–Schwarz inequality

$$\left(\sum_i m_i\right)^2 \leq \sum_i 1 \cdot \sum_i m_i^2.$$

(b) If $M = \left(\begin{smallmatrix} 1 & 1 \\ -1 & 0 \end{smallmatrix}\right)$, then $(\operatorname{Tr} M)^2 = 1$, but $M^2 = \left(\begin{smallmatrix} 0 & 1 \\ -1 & -1 \end{smallmatrix}\right)$, and $\operatorname{Tr}(M^2) = -1$.

4. Multiply $0 = (\rho v^\beta)_{;\beta} v^\alpha + \rho v^\beta (v^\alpha{}_{;\beta})$ by $v_\alpha$. From $\langle v | v \rangle = -1$ it follows that $v_\alpha (v^\alpha{}_{;\beta}) = 0$, and we conclude that $0 = (\rho v^\beta)_{;\beta}$. In that case, $v^\beta (v^\alpha{}_{;\beta}) = \langle e^\alpha | D_v v \rangle = 0$. The equivalence of the expressions for $c$ follows from

$$L_v {}^*\mathbf{1} = di_v {}^*\mathbf{1} = d{}^*v,$$

$$\begin{aligned} d({}^*e^\alpha v_\alpha) &= -v_\alpha \omega^\alpha{}_\beta \wedge {}^*e^\beta + dv_\beta \wedge {}^*e^\beta = {}^*\langle e^\beta | dv_\beta - v_\alpha \omega^\alpha{}_\beta \rangle \\ &= {}^*\langle e^\beta | D_{e_\beta} v \rangle = {}^*(v^\beta{}_{;\beta}). \end{aligned}$$

5. Let $d(\lambda) = \inf_{C^1 \supset U \ni \lambda} \sup_{\bar{\lambda} \in U} d(\bar{\lambda})$. This is upper semicontinuous and workable as long as the supremum is finite for sufficiently small $U$, which is the case as a corollary of the proof of (4.6.21) (see Problem 6(c)). If $f\colon \mathbb{R}\backslash\{0\} \to \mathbb{R}$ sends $x \to |x|^{-1}$, then this is not the case at $\{0\}$, and this function can not be extended upper semicontinuously to a function $f\colon \mathbb{R} \to \mathbb{R}$. Incidentally, the above extension is maximally continuous; for example, to $f\colon \mathbb{R}\backslash\{0\} \to \mathbb{R}$, $x \to |x|$ it ascribes the value $f(0) = 0$, whereas $f(0) = a > 0$ would make the extension only upper semicontinuous.

6. (a) If $J^-(x) \cap D^+(S)$ were not compact, then there would exist an infinite, locally finite covering with relatively compact neighborhoods $U_i$ with $a_i \in U_i$ for $\{a_i\}$ without a point of accumulation. Let $x \in U_1$ and $\gamma_i$ be a family of causal curves from $x$ to $a_i$. Then $\gamma_i \cap \partial U_1$ has a point of accumulation $h_1$. If $c_1$ is a causal curve from $x$ to $h_1$, then $c_1$ contains a point $x_1$ that lies not only in $U_1$ but also in another set $U_2$. Since $J^-(x_1)$ contains a neighborhood of $h_1$ by Assumption (4.6.20(b)), it also contains an infinite subfamily $\{\gamma_{1_i}\}$ of the $\{\gamma_i\}$, and consequently infinitely many $a_{1_i}$. There is a point of accumulation $h_2$ for $\{\gamma_{1_i}\} \cap \partial U_2 \cap J^-(x_1)$, and there exists a causal curve $c_2$ from $x_1$ to $h_2$, and so on (Figure 70). This procedure yields a causal line connecting $x, x_1, x_2, \ldots$, which can not be extended farther downwards, since the $a_i$ have no point of accumulation. However, for the same reason, it can not intersect $S$, as otherwise one of the relatively compact $U_i$ would contain infinitely many $a_j$. The existence of a nonextensible causal curve not meeting $S$ contradicts the definition of $D^+(S)$, and therefore $J^-(x) \cap D^+(S)$ must be compact.

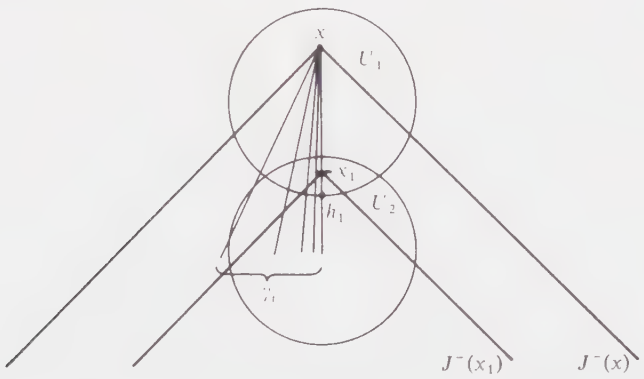



Figure 70 Construction of a nonextensible, causal curve that does not intersect $S$

(b) $C(x, S)$ is compact as a metric space if it is complete and precompact. A uniform limit of causal curves is continuous [(22], 7.2.1), and, because of (4.6.20(b)), causal. Therefore completeness follows immediately. Precompact means that for all $\varepsilon$ there exists a finite covering of $C(x, S)$ with neighborhoods of diameter $\varepsilon$.

Such a neighborhood of a curve $\gamma$ is the set of curves

$$\{\gamma' : \sup_{y \in \gamma'} \inf_{x \in \gamma} \rho(x, y) < \varepsilon\},$$

where $\rho$ is a distance function for the metric on $M$. Since $J^-(x) \cap D^+(S)$ is compact, we can cover it with finitely many $A_i$ = the interior of $J^-(a_i) \cap J^+(a_i')$, $i = 1, \ldots, n$, with diameter $< \varepsilon$. Let $x$ be in $A_1$, and form all unions

$$B_j = \bigcup_{k=1}^{n_j} A_{jk},$$

with

$$A_{j_1} = A_1, \qquad A_{j_{n_j}} \cap S \neq \emptyset,$$
$$A_{j_k} \cap A_{j_{k+1}} \neq \emptyset, \qquad J^-(A_{j_k}) \cap A_{j_{k+2}} \neq \emptyset.$$

The $B_j' = \{\gamma \in C(x, S) : \gamma \subset B_j\}$ are a covering for $C(x, S)$, since every causal curve of $x$ must be in some such union. The number of the $B_j$ is finite, and their diameter $< \varepsilon$.

(c) We need to show that for all $\varepsilon$ there exists a neighborhood $U$ of $\lambda$ such that $d(\bar{\lambda}) < d(\lambda) + \varepsilon$ for all $\bar{\lambda} \in U$. For this purpose, we use co-moving coordinates moving with $\lambda$, so that $\lambda$ is one of the time axes $x^j$ = constant, $j = 1, 2, 3$, and the $g_{0j}$ vanish. Since $d$ does not depend on the choice of the curve parameter, we can take this parameter as $x^0$; we then have to compare

$$d(\lambda) = \int_{\lambda} dx^0 \sqrt{-g_{00}}$$

with

$$d(\bar{\lambda}) = \int_{\bar{\lambda}} dx^0 \sqrt{-g_{00} - g_{jk} \dot{\bar{\lambda}}^j \dot{\bar{\lambda}}^k}.$$

But since $g_{jk} \dot{\bar{\lambda}}^j \dot{\bar{\lambda}}^k > 0$ and $g_{00}$ is uniformly continuous, as a continuous function on the compact set $J^-(x) \cap D^+(S)$, we can make $d(\bar{\lambda}) < d(\lambda) + \varepsilon$ by taking a small enough neighborhood.

## Some Difficult Problems

1. Only approximate solutions are known for the diffraction at a slit. Find bounds for the errors.
2. Give a general, rigorous formulation of Babinet's principle [2].
3. Show that the Green function for diffraction at a wedge has causal support properties.
4. Find a solution of Einstein's equations that describes the emission of gravitational waves.
5. Harmonic coordinates are used to prove that Einstein's equations are hyperbolic. Give a purely geometric formulation of this state of affairs, without reference to particular coordinates.
6. Discover singularity theorems that show that $M$ is not only geodesically incomplete, but that the curvature invariants are in fact unbounded, in the right circumstances.
7. Solve the general relativistic two-body problem.

# Bibliography

## Works Cited in the Text


[1] J. D. Jackson. Classical Electrodynamics. New York: Wiley, 1975.

[2] W. Franz. Theorie der Beugung elektromagnetischer Wellen, Ergebnisse der angew. Math., vol. 4. Berlin: Springer, 1957.

[3] T. Fulton and F. Rohrlich. Classical Radiation from a Uniformly Accelerated Charge. *Ann. Phys.* **9**, 499–517, 1960.

[4] K. Yano. The Theory of Lie Derivatives and Its Applications. Amsterdam: North-Holland, 1955.

[5] H. M. Nussenzveig. High Frequency Scattering by an Impenetrable Sphere. *Ann. Phys.* **34**, 23–95, 1965.

[6] F. London. Superfluids, vol. I: Macroscopic Theory of Superconductivity. New York: Wiley, 1950.

[7] J. L. Anderson. Principles of Relativity Physics. New York: Academic Press, 1967.

[8] R. U. Sexl, H. K. Urbantke. Gravitation und Kosmologie, BI-Hochschultaschenbuch, Mannheim: BI-Wissenschaftsverlag, 1974.

[9] C. W. Misner, K. S. Thorne, J. A. Wheeler. Gravitation. San Francisco: Freeman, 1973.

[10] S. Weinberg. Gravitation and Cosmology. New York: Wiley, 1972.

[11] J. M. Souriau. Géometrie et Rélativité. Paris: Hermann, 1964.

[12] M. D. Kruskal. Maximal Extension of Schwarzschild Metric. *Phys. Rev.* **119**, 1743–1745, 1960.

[13] J. C. Graves and D. R. Brill. Oscillatory Character of Reissner–Nordström Metric for an Ideal Charged Wormhole. *Phys. Rev.* **120**, 1507–1513, 1960.

[14] B. Carter. The Complete Analytic Extension of the Reissner–Nordström Metric in the Special Case $e^2 = m^2$. *Phys. Letters* **21**, 423–424, 1966.

[15] A. Trautman. Theory of Gravitation. In: The Physicist's Conception of Nature, J. Mehra, ed. Boston: D. Reidel, 1973.

[16] C. N. Yang. Integral Formalism for Gauge Fields. *Phys. Rev. Lett.* **33**, 445–447, 1974.
[17] W. Rühl. Finite Conformal Transformations in Local Quantum Field Theory, in: Electromagnetic Interactions and Field Theory, *Acta Phys. Austriaca Suppl.* **XIV**, 643-646, 1975.
[18] M. Schönberg. *Revista Brasileira de Fisica 1*, 91, 1971.
[19] A. Uhlmann. *Wissenschaftliche Zeitschrift der Friedrich-Schiller-Universität* **8**, 31, 1958.
[20] M. Abramowitz and I. E. Stegun, eds. Handbook of Mathematical Functions, Applied Mathematics Series 55. Washington: National Bureau of Standards, 1964.
[21] F. Hehl, P. von der Heyde, and G. D. Kerlick. General Relativity with Spin and Torsion: Foundations and Prospects. *Rev. Mod. Phys.* **48**, 393–416, 1976.
[22] J. Dieudonné. Foundations of Modern Analysis, vols. III and IV. New York: Academic Press, 1972 and 1974.
[23] A. Schild. Electromagnetic Two-Body Problem. *Phys. Rev.* **131**, 2762–2766, 1962.
[24] F. Rohrlich. Classical Charged Particles: Foundations of Their Theory. Reading, Mass.: Addison-Wesley, 1965.
[25] L. P. Eisenhart. Riemannian Geometry. Princeton: Princeton Univ. Press, 1949.
[26] V. Fock. The Theory of Space, Time and Gravitation. New York: Macmillan, 1964.
[27] A. Lichnerowicz. Théories Rélativistes de la Gravitation et de l'Electromagnetisme: Rélativité Générale et Théories Unitaires. Paris: Masson, 1955.
[28] J. Wess and B. Zumino. Superspace Formulation of Supergravity. *Phys. Letters* **66B**, 361–364, 1977.
[29] A. Trautman. Conservation Laws in General Relativity. In: Gravitation, an Introduction to Current Research, L. Witten, ed. New York: Wiley, 1962.
[30] E. Pechlaner and R. Sexl. On Quadratic Lagrangians in General Relativity. *Comm. Math. Phys.* **2**, 165–175, 1966.
[31] F. Hoyle and J. V. Narlikar. Cosmological Models in Conformally Invariant Gravitational Theory—II. A New Model. *Mon. Not. Roy. Astr. Soc.* **155**, 323–335, 1972.
[32] L. Landau and E. Lifshitz. The Classical Theory of Fields. Reading, Mass.: Addison-Wesley, 1977.
[33] G. F. R. Ellis and S. W. Hawking. The Large Scale Structure of Space-Time. Cambridge: At the University Press, 1973.
[34] R. P. Geroch. What is a Singularity in General Relativity? *Ann. Phys.* **48**, 526–540, 1968.
[35] M. Fierz and R. Jost. Affine Vollständigkeit und kompakte Lorentz'sche Mannigfaltigkeiten. *Helv. Phys. Acta* **38**, 137–141, 1965.
[36] F. Hoyle et al. In: Quasi-stellar Sources and Gravitational Collapse, I. Robinson, ed. Chicago: The Univ. of Chicago Press, 1965.
[37] M. Reed and B. Simon. Methods of Modern Mathematical Physics, in four volumes. New York: Academic Press, 1974–1979.
[38] H. Bondi and T. Gold. The Field of a Uniformly Accelerated Charge, with Special Reference to the Problem of Gravitational Acceleration. *Proc. Roy. Soc. London* **A229** 416–424, 1955.
[39] W. Thirring and R. Wallner. The Use of Exterior Forms in Einstein's Gravitation Theory. *Revista Brasileira de Fisica*, to appear.

## Further Reading

*1. Alternating Differential Forms*

H. Cartan. Differential Calculus. Paris: Hermann, 1971.
H. Cartan. Differential Forms. Paris: Hermann, 1970.
G. A. Deschamps. Exterior Differential Forms. In: Mathematics Applied to Physics, E. Roubine, ed. New York: Springer, 1970.
H. Flanders. Differential Forms with Applications to the Physical Sciences. New York: Academic Press, 1963.
S. J. Goldberg. Curvature and Homology. New York: Academic Press, 1962.
W. Greub, S. Halperin, and R. Vanstone. Connections, Curvature, and Cohomology. New York: Academic Press, 1972.
H. Holmann and H. Rummler. Alternierende Differentialformen. Mannheim: BI-Wissenschaftsverlag, 1972.

*2. Tensor Analysis and Geometry of Manifolds*

L. Auslander and R. E. MacKenzie Introduction to Differential Manifolds. New York: McGraw-Hill, 1963.
R. L. Bishop and R. J. Crittenden. Geometry on Manifolds. New York: Academic Press, 1964.
R. L. Bishop and S. I. Goldberg. Tensor Analysis on Manifolds. New York: Macmillan, 1968.
F. Brickell and R. S. Clark. Differentiable Manifolds. New York: Van Nostrand-Reinhold, 1970.
T. Bröcker and K. Jänich. Einführung in die Differentialtopologie, Heidelberger Taschenbuch 143. Heidelberg: Springer, 1968.
Y. Choquet-Bruhat, C. DeWitt-Morette, and M. Dillard-Bleick. Analysis, Manifolds, and Physics. Amsterdam: North-Holland, 1978.
D. Gromoll, W. Klingenberg, and W. Meyer. Riemannsche Geometrie im Großen, Lecture Notes in Mathematics, 55. New York: Springer, 1968.
N. J. Hicks. Notes on Differential Geometry. New York: Van Nostrand-Reinhold, 1971.
S. Kobayashi and K. Nomizu. Foundations of Differential Geometry, vols. I and II. New York: Interscience, 1963 and 1969.
A. Lichnerowicz. Elements of Tensor Analysis. New York: Wiley, 1962.
C. W. Misner. Differential Geometry. In: Relativity, Groups, and Topology, C. DeWitt and B. S. DeWitt, eds. New York: Gordon and Breach, 1964.
E. Nelson. Tensor Analysis. Princeton: Princeton Univ. Press, 1967.
S. Sternberg. Lectures on Differential Geometry. Englewood Cliffs, N.J.: Prentice-Hall, 1964.
T. J. Willmore. An Introduction to Differential Geometry. Oxford: Oxford Univ. Press, 1959.
J. A. Wolf. Spaces of Constant Curvature. New York: McGraw-Hill, 1967.

*3. General Relativity*

R. Adler, M. Bazin, and M. Schiffer. Introduction to General Relativity. New York: McGraw-Hill, 1965.

A. Einstein. The Meaning of Relativity. Princeton: Princeton Univ. Press, 1955.
W. Pauli. Theory of Relativity. New York: Pergamon, 1958.
W. Rindler. Essential Relativity. New York: Springer, 1977.
R. U. Sexl and H. K. Urbantke. Gravitation und Kosmologie. BI-Hochschultachsenbuch. Mannheim: BI-Wissenschaftsverlag, 1975.
J. L. Synge. Relativity, the General Theory. Amsterdam: North-Holland, 1965.
A. Trautman, F. Pirani, and H. Bondi. Lectures on General Relativity. Englewood Cliffs, N.J.: Prentice-Hall, 1972.

*4. Global Analysis*

Y. Choquet-Bruhat and R. Geroch.Global Aspects of the Cauchy Problem in General Relativity. *Comm. Math. Phys.* **14**, 329–335, 1969.
G. F. R. Ellis and D. W. Sciama. Global and Nonglobal Problems in Cosmology. In: General Relativity, Papers in Honor of J. L. Synge, L. O'Raifeartaigh, ed. Oxford: The Clarendon Press, 1972.
D. Farnsworth, J. Fink, J. Porter, and A. Thomson, eds. Methods of Local and Global Differential Geometry in General Relativity, Lecture Notes in Physics 14. New York: Springer, 1972.
R. P. Geroch. Topology in General Relativity. *J. Math. Phys.* **8**, 782–786, 1967.
R. P. Geroch. Domain of Dependence. *J. Math. Phys.* **11**, 437–449, 1970.
R. P. Geroch Space-Time Structure from a Global Point of View. In: General Relativity and Cosmology, R. K. Sachs, ed. New York: Academic Press, 1971.
ICTP, Global Analysis and its Applications, vols. I, II, and III. Lectures Presented at an International Seminar Course at Trieste from 4 July to 25 August, 1972. New York: Unipub, 1975.
W. Kundt. Global Theory of Spacetime. In: Differential Topology, Differential Geometry and Applications, J. R. Vanstone, ed. Montreal: Canadian Mathematical Congress, 1972.
A. Lichnerowicz. Topics on Space-Time. n: Batelle Rencontres: 1967 Lectures in Mathematics and Physics, C. DeWitt and J. A. Wheeler, eds. New York: Benjamin, 1968.
R. Penrose. Structure of Space-Time. *Ibid.*

*5. Proceedings, Summer Schools, and Collected Papers*

P. G. Bergmann, E. J. Fenyves, and L. Motz, eds. Seventh Texas Symposium on Relativistic Astrophysics. *Annals of the New York Acad. of Sci.* **262**, 1975.
M. Carmeli, S. Fickler, and L. Witten, eds. Relativity. New York: Plenum, 1970.
H-Y. Chiu and W. F. Hoffman, eds. Gravitation and Relativity New York: Benjamin, 1964.
C. DeWitt and J. A. Wheeler, eds. Batelle Rencontres: 1967 Lectures in Mathematics and Breach, 1964.
C. DeWitt and J. A. Wheeler, eds. Batelle Rencontres: 1967 Lectures in Mathematics and Physics. New York: Benjamin, 1968.
Editorial Committee. Recent Developments in General Relativity. New York: Macmillan, 1962.
J. Ehlers, ed. Relativity Theory and Astrophysics. Providence: Amer. Math. Soc., 1967.
W. Israel, ed. Relativity, Astrophysics, and Cosmology. Boston: D. Reidel, 1973.

C. W. Kilmister, ed. General Theory of Relativity, Selected Readings in Physics. New York: Pergamon, 1973.
C. G. Kuper and A. Peres, eds. Relativity and Gravitation. New York: Gordon and Breach, 1971.
L. O'Raifeartaigh, ed. General Relativity, Papers in Honor of J. L. Synge. Oxford: The Clarendon Press, 1972.
R. K. Sachs, ed. General Relativity and Cosmology, Proceedings of Course 47 of the International School of Physics "Enrico Fermi." New York: Academic Press, 1971.
G. Shaviv and J. Rosen, eds. General Relativity and Gravitation. New York: Wiley, 1975.
P. Suppes, ed. Space, Time, and Geometry. Boston: D. Reidel, 1973.
J. R. Vanstone, ed. Differential Topology, Differential Geometry and Applications, Proceedings of the Thirteenth Biennial Seminar of the Canadian Mathematical Congress. Montreal: Canadian Mathematical Congress, 1972.
L. Witten, ed. Gravitation, an Introduction to Current Research. New York: Wiley, 1962.

*Section 4.1*

R. L. Bishop and S. I. Goldberg. Tensor Analysis on Manifolds. New York: Macmillan, 1968.
N. J. Hicks. Notes on Differential Geometry. New York: Van Nostrand-Reinhold, 1971.
S. Kobayashi and K. Nomizu. Foundations of Differential Geometry, vols. I and II. New York: Interscience, 1963 and 1969.
B. Schmidt. Differential Geometry from a Modern Standpoint. In: Relativity, Astrophysics, and Cosmology, W. Israel, ed. Boston: D. Reidel, 1973.

*Section 4.2*

J. Gel'fand and S. Fomin. Calculus of Variations. Englewood Cliffs, N.J.: Prentice-Hall, 1963.
P. Havas. On Theories of Gravitation with Higher Order Field Equations, *Gen. Rel. Grav.* **8**, 631, 1977.
D. Lovelock and H. Rund. Variational Principles in the General Theory of Relativity. *Jahresbericht der Deutschen Mathematiker-Vereinigung*, **74**, No. 1/2, 1972.
A. Trautman. Conservation Laws in General Relativity. In: Gravitation, an Introduction to Current Research, L. Witten, ed. New York: Wiley, 1962.

*Sections 4.3 and 4.4*

S. Helgason. Lie Groups and Symmetric Spaces. In: Batelle Rencontres: 1967 Lectures in Mathematics and Physics, C. DeWitt and J. A. Wheeler, eds. New York: Benjamin, 1968.

*Section 4.5*

B. K. Harrison, K. S. Thorne, M. Wakano, and J. A. Wheeler. Gravitational Theory and Gravitational Collapse. Chicago: The Univ. of Chicago Press, 1965.

H. Scheffler and H. Elsässer, Physik der Sterne und der Sonne. Mannheim: BI-Wissenschaftsverlag, 1974.
Ya. B. Zel'dovich and I. D. Novikov. Relativistic Astrophysics, vols. I and II. Chicago: The Univ. of Chicago Press, 1971.

*Section 4.6*

C. J. S. Clarke. The Classification of Singularities. *Gen. Rel. Grav.* **6**, 35–40, 1975.
C. J. S. Clarke. Space-Time Singularities. *Comm. Math. Phys.* **49**, 17–23, 1976.
G. F. R. Ellis and B. Schmidt. Singular Space-Times. *Gen. Rel. Grav.* **8**, 915–953, 1977.
R. P. Geroch. Singularities in the Spacetime of General Relativity. Their Definition, Existence, and Local Characterization. Dissertation, Princeton University, 1967.
R. P. Geroch, What is a Singularity in General Relativity? *Ann. Phys.* **48**, 526–540, 1968.
S. W. Hawking. Singularities and the Geometry of Spacetime. Essay submitted for the Adams Prize, Cambridge, 1966.
S. W. Hawking. The Occurrence of Singularities in Cosmology I, II, III. *Proc. Roy. Soc. London* **294A**, 511–521, 1966. *Ibid.* **295A**, 490–493, 1966. *Ibid.* **300A**, 187–201, 1967.
W. Kundt. Recent Progress in Cosmology. Springer Tracts in Modern Physics, 47. New York: Springer, 1968.
R. Penrose. Gravitational Collapse and Space-Time Singularities. *Phys. Rev. Lett.* **14**, 57–59, 1965.

# Index


Action 46
ADM energy 181
Affine connection 16
Anti-de Sitter universe 200
Asymptotic fields 66

Basis 10
Bianchi identity 164, 188
Birkhoff's theorem 212
Black hole 8

Cartan's structure equation 162
Cauchy surface 19
Causal curve 53
Characteristics 17, 58, 185
Christoffel symbol 168
Closed form 12
Codifferential 15
Conformal transformation 52
Convergence of the flow 239
Cornu's spiral 135
Cosmic censorship 219
Cosmological principle 197
Cosmological term 179, 202
Coulomb potential 67
Covariant differentiation 155
Covariant Lie derivative 160
Current density 33
Curvature form 162

De Sitter metric 205
De Sitter universe 198
Delta function 20
Diffraction 125
Dipole radiation 6
Discontinuity surface 20, 185
Distribution 20
Domain of influence 53
Double star 8
Duality map 14

Einstein's equations 41, 181
Electromagnetic radiation 1, 3
Electromagnetic waves 2
Energy momentum current 36
Energy momentum form 81, 108
Energy momentum loss by radiation 68, 95
Energy momentum tensor 48, 108
Equivalence principle 39
Event horizon 199
Exact form 12
Exterior differential 12
Exterior product 11

Fibre metric 156
Fictitious force 39
Field-strength form 29
Flat 163
Form 11
Frame 11

Frauenhofer region 150
Fresnel's integral 134, 150
Friedmann universe 205
Future 342

Gauge group 157
Gauge theory 174
Gauge transformation 174
Gauss's theorem 13
Geodesic deviation 239
Geodesic vector field 169
Geodesically complete 237
Geometric optics 129
Gravitational radiation 7
Gravitational wave 208
Green's formula 22
Green function 22
Group velocity 119

Harmonic coordinates 178, 186
Heaviside step function 20
Helmholtz's circulation theorem 103
Hertz's dipole 84
Homogeneous 193
Homogeneous star 226
Horizon 199
Hyperbolic motion 77
Hypersurface 19

Integral 13
Interior product 14
Isotropic 193

Killing vector field 50, 169
Kirchhoff's theory of diffraction 139

Lagrangian 47
Landau–Lifschitz form 183
Laplace–Beltrami operator 15
Larmor's formula 4, 91
Length of a curve 243
Lie derivative 16
Liénard–Wiechert potential 66
Lightlike coordinates 22
Linear approximation 186
Local Lorentz transformation 177, 178
London's equation 103
Lorentz force 37
Lorentz gauge 62

Magnetic charge 30
Mass renormalization 97
Maximally symmetric space 191
Maxwell's equation 29
Metallic boundary conditions 109
Minimal frequency 122

Naked singularity 219
Natural basis 12
Neutron star 8
Noether's theorem 48
Normal mode 118

Oppenheimer–Snyder solution 232
Orientable 14
Orthogonal basis 16

Parallel transport 158, 161
Partial differential equation 17
Particle horizon 199
Past 243
Penrose diagram 199
Perfect cosmological principle 197
*p*-form valued sections 157
Plasma frequency 105
Poincaré transformation 34
Point particle 38
Poynting's vector 52
Principle of equivalence 40
Proper time 243

Quasiregular singularity 248

Radiation field 67, 94
Radiative reaction 68
Reissner–Nordstrøm metric 205
Renormalized equations of motion 97
Resonant cavity 121
Restriction 12
Retarded Green function 60
Riemann–Christoffel tensor 171
Riemann structure 13
Robertson–Walker metric 204
Rotating basis 39
Rotating charge 80
Run-away solution 99

Saddle-point method 131
Scattering angle 151
Scattering cross-section 5, 151, 152

Schwarzschild metric 215
Section 155
Shadow 135, 142, 148
Signal velocity 119
Soldering form 166
Static 193
Stationary 193
Steepest descent 131
Step function 20
Stokes's theorem 13
Structure equation 162
Superconductor 102
Supernova 224
Surface area of the $m$-sphere 204
Surface of discontinuity 185
Synchrotron radiation 92

Tachyon 121
Tensor field 10
Tensor product 11
Tetrad 11
TE and TM solutions 123
Tidal force 248
Tolman–Oppenheimer–Volkoff equation 225
Torsion 166
Total charge 35

Uniform acceleration 77
Uniform motion 75

Vector field 10
Vector potential 31

Wave front 120
Wave guide 116
Wedge product 11
Weyl forms 171

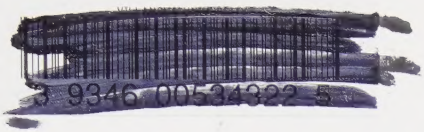
Yang–Mills theory 175, 177
Yukawa potential 105

DATE DUE

MAY 20 '87
DISCHARGED

SEP 23 '89
DISCHARGED

DEMCO NO. 38-298


3 9346 00534322 5

The second edition of this well-received introduction to classical field theory contains a new section on gauge theories and gravitation, in addition to numerous minor changes in the text, including several new figures.

ISBN 0-387-96266-2 Springer-Verlag New York
ISBN 3-211-96266-2 Springer-Verlag Wien